

Máster Interuniversitario en Estadística e Investigación Operativa UPC-UB

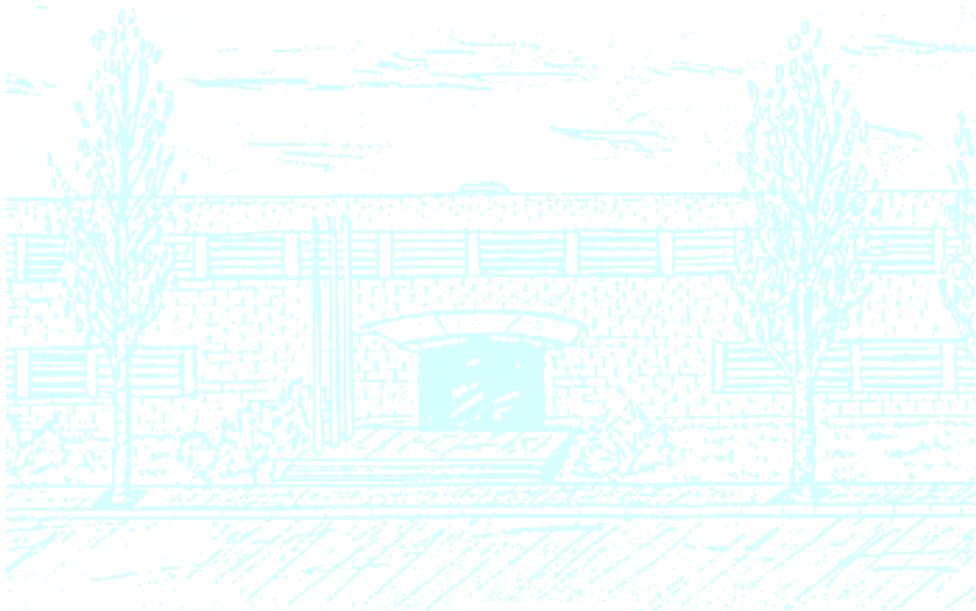
Título: Un método basado en la programación matemática para seleccionar líneas de autobuses candidatas en extensiones de sistemas de transporte público

Autor: Grace Kelly Maureira Alegría

Director: Esteve Codina Sancho

Departamento: Estadística e Investigación Operativa

Universidad: UPC - UB



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística



UNIVERSITAT DE BARCELONA

Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Tesis de máster

**Un método basado en la programación
matemática para seleccionar líneas de
autobuses candidatas en extensiones de
sistemas de transporte público**

Grace Kelly Maureira Alegría

Director: Esteve Codina Sancho

Departamento de Estadística e Investigación Operativa

Agradecimientos

En primer lugar, quiero agradecer a mi marido Francisco por su apoyo incondicional en este proceso.

También quiero agradecer al profesor Esteve por haber confiado en mi para el desarrollo de este trabajo.

Por último, agradezco a CONICYT - Chile por haberme otorgado una de las Becas de Magister en el Extranjero (Becas Chile 2016) que me permitió estudiar este máster.

Resumen

Palabras clave: Transporte público, Establecimiento de frecuencias, Selección líneas de autobuses, Programación matemática

MSC2000: 90B06, 90B20, 90C11, 90C59, 90C90

En este trabajo, un modelo de programación matemática para la selección de líneas candidatas en extensiones del sistema público es propuesto a partir de la linealización del modelo presentado por Codina-Marín-Lopez (2013) y su aplicación a casos en situaciones de mediana congestión. Así mismo, se proponen diferentes métodos de resolución basados en el método del Subgradiente y el algoritmo de Cutting Plane de Dantzig.

Al método del Subgradiente se le aplican dos métodos para el cálculo de las longitudes de paso para la actualización de los multiplicadores de Lagrange: el método de medias ponderadas sucesivas (MSWA) y el método de medias ponderadas autorreguladas. Al algoritmo de Cutting Plane se le aplican dos diferentes métodos para acelerar su convergencia: la actualización de los multiplicadores de Lagrange a través del MSWA y la reducción de cortes cada cierta cantidad de iteraciones.

El modelo propuesto minimiza los costos de implementación y operación del servicio más el valor económico del tiempo total de viaje de los pasajeros. Las soluciones que entrega son (a) el número de autobuses y servicios a asignar a cada línea candidata, (b) las rutas que debe seguir el pasajero entre cada par origen-destino, y (c) el tiempo de espera de los pasajeros en las paradas.

Los métodos de resolución propuestos se han implementado en AMPL y resuelto con CPLEX. Se ha hecho una comparación del funcionamiento de estos métodos a partir de los ejemplos numéricos proporcionados por dos redes auxiliares del transporte público en situaciones de interrupción del servicio: una red de tamaño pequeño que corresponde a la interrupción de un corredor ferroviario en Madrid, y una red de tamaño mediano situada en Barcelona durante una interrupción de la línea 1 del metro.

Los resultados obtenidos muestran que únicamente el algoritmo de Cutting Plane resuelve el modelo. Sin embargo, solo la aplicación del método MSWA lo acelera.

Abstract

Keywords: Public transportation, Frequency setting, Bus lines selection, Mathematical programming

MSC2000: 200090B06, 90B20, 90C11, 90C59, 90C90

In this paper, a mathematical programming model for selecting candidate bus lines in extensions of public transportation is proposed based on the linearization of the model presented by Codina-Marín-Lopez (2013) and its application to cases in situations of medium congestion. Furthermore, different solution methods are proposed based on the Subgradient method and Dantzig's Cutting Plane algorithm.

Two methods are used to calculate the step lengths for updating the Lagrange multipliers: the method of successive weighted averages (MSWA) and the self-regulated averaging method. Two different methods are applied to the Cutting Plane algorithm to accelerate its convergence: the updating of the Lagrange multipliers through the MSWA and the reduction of cuts every certain number of iterations.

The proposed model minimizes the total set-up/operational costs of the extension service plus the economic value of the total user travel time. The output of the model is as follows: (a) the number of bus units and services to assign to each of the candidate line; (b) the routes to be followed by users passengers of each origin-destination pairs; (c) the passenger waiting time for boarding at a line.

The proposed solution methods have been implemented in AMPL and solved using CPLEX. A comparison of the performance of these methods has been made from the numerical examples provided by two auxiliary public transportation networks during disruptions: a small network that corresponds to the disruption of a railway corridor in Madrid, and a medium-sized network located in Barcelona during a disruption on the metro line 1.

The results obtained show that only the Cutting Plane algorithm solves the model. However, the application of the MSWA is the only one which accelerates it.

Notación

N	Conjunto de todos los nodos de la red
N_G	Conjunto de nodos que representan la red de transbordos
\hat{N}_G	Conjunto de nodos que actúan como paradas, $\hat{N}_G \subseteq N_G$
A	Conjunto de todos los arcos de la red
A_G	Conjunto de arcos que representan la red de transbordos (en tierra)
A_b	Conjunto de arcos cuyos nodos no están en N_G y que acceden a la estación $b \in \hat{N}_G$
A_ℓ	Conjunto de arcos cuyos nodos no están en N_G y modelizan la línea $\ell \in L$
$A_{b,\ell}$	Conjunto de arcos $a \in A_b$ correspondientes a la línea $\ell \in L$, $A_{b,\ell} = A_b \cap A_\ell$
L	Conjunto de líneas de autobuses
L_b	Conjunto de líneas que contienen la estación $b \in \hat{N}_G$, $L_b = \{\ell \in L A_{b,\ell} \neq \emptyset\}$
Π_ℓ	Conjunto de paradas que definen una línea $\ell \in L$
W	Conjunto de pares de nodos origen-destino, $W \in N_G$
D	Conjunto de nodos destino
b	Índice de parada de autobús, $b \in \hat{N}_G$
ℓ	Índice de línea, $\ell \in L$
ω	Índice de par origen-destino, $\omega = (p, q) \in W$
g_ω	Flujo de pasajeros en cada par origen-destino ω
\mathbf{p}	Unidades de autobuses disponibles
c^ℓ	Capacidad de pasajeros de la línea $\ell \in L$
κ_b	Tiempo medio de servicio en la parada $b \in \hat{N}_G$
s_b	Máximo número de áreas de carga en la parada $b \in \hat{N}_G$
H	Período de tiempo
C_ℓ	Cantidad de ciclos por línea $\ell \in L$
C_ℓ^0	Tiempo total de viaje libre entre dos paradas consecutivas de la línea $\ell \in L$
t_a^0	Tiempo de viaje libre entre dos paradas consecutivas de la línea $\ell \in L$
w_b^0	Tiempo medio de espera en el espacio L0 en la parada $b \in \hat{N}_G$
w_b^1	Tiempo medio de espera en el espacio L1 en la parada $b \in \hat{N}_G$
\mathcal{L}_b^0	Máximo número de buses permitidos en el espacio L0 en la parada $b \in \hat{N}_G$
\mathcal{L}_b^1	Máximo número de buses permitidos en el espacio L1 en la parada $b \in \hat{N}_G$
η_b^0	Factor de ocupación al 95 % en el espacio L0 en la parada $b \in \hat{N}_G$
η_b^1	Factor de ocupación al 95 % en el espacio L1 en la parada $b \in \hat{N}_G$
\hat{Z}_b	Máximo número de servicios en la parada $b \in \hat{N}_G$

h_{max}	Máximo headway
\bar{f}	Cota superior de la frecuencia de servicios
\underline{f}	Cota inferior de la frecuencia de servicios
χ_a	Coefficiente de variación del headway en el arco $a = a(\ell, b), b \in \hat{N}_G, \ell \in L$
η_b	Razón entre la ocupación máxima y media al 95 % en la parada $b \in \hat{N}_G$
\hat{N}_G	Máximo número de pasajeros que pueden esperar en la parada $b \in \hat{N}_G$
T_a	Tiempo de viaje en el arco $a = (i, j), i \in N_G, j \notin N_G$
ζ^ℓ	Costo de alquilar un autobús para la línea $\ell \in L$ [€/u. autobús]
γ^ℓ	Costos de operación de un servicio de la $\ell \in L$ [€/servicio]
θ	Valor económico del tiempo total de viaje de los pasajeros [€/u. de tiempo.]
n^ℓ	Número de autobuses asignados a la línea $\ell \in L$
z^ℓ	Número de servicios asignados a la línea $\ell \in L$
λ^ℓ	$\lambda^\ell = 1$ si se asignan servicios a la línea $\ell \in L$, $\lambda^\ell = 0$ si no
v_a	Flujo total de pasajeros en el arco $a \in A$
ζ_a	Tiempo total que los pasajeros esperan en una parada para abordar una línea, $a = a(\ell, b), b \in \Pi_\ell, \ell \in L$

Índice general

Índice de figuras	v
Índice de cuadros	vii
Capítulo 1. Introducción	1
1.1. Revisión literaria	2
1.1.1. Notación	2
1.1.2. Modelos	3
1.1.2.1. Modelo de Ceder y Wilson (1986)	3
1.1.2.2. Modelo de Baaaj y Mahmassani (1991)	4
1.1.2.3. Modelo de Israeli y Ceder (1993)	4
1.1.2.4. Modelo de Tom y Mohan (2003)	5
1.1.3. Principales características de los modelos para el TNDP	5
1.2. Objetivos	5
1.3. Organización del documento	6
Capítulo 2. Métodos de descomposición	7
2.1. Programación Lineal	7
2.2. Relajación Lagrangiana	8
2.3. Método del Subgradiente	9
2.3.1. Longitud de paso	10
2.3.2. Criterio de parada	11
2.3.3. Recuperación de la solución primal	11
2.4. Algoritmo de Cutting Plane de Dantzig	12
2.4.1. Criterio de parada	12
2.4.2. Obtención de una solución primal factible	13
2.4.3. Aceleración del algoritmo de Cutting Plane	13
Capítulo 3. Definición del problema	15
3.1. Modelo de red	15
3.2. Flujo de pasajeros	17
3.3. Modelado de los tiempos operacionales	17
3.3.1. Modelado de las paradas de autobús	17
3.3.1.1. Tiempos de servicio de autobús y de pasajeros en las paradas	18
3.3.1.2. Ciclos de línea y cola de autobuses en las paradas	18
3.3.1.3. Máximo flujo de entrada de autobuses en una parada	19

3.3.1.4. Limitaciones en el máximo headway y frecuencias admisibles en las líneas	20
3.3.2. Tiempo total de espera de los pasajeros en la parada	20
3.3.3. Modelado de capacidad finita para pasajeros esperando en una parada de autobús	22
3.4. Problema de programación matemática	23
Capítulo 4. Aplicación de los métodos de descomposición	25
4.1. Modelo relajado	25
4.2. Resolución de $M0_{LR}$ mediante el método del Subgradiente	26
4.3. Resolución de $M0_{LR}$ mediante el algoritmo de Cutting Plane	27
4.3.1. Obtención de una solución factible inicial	28
4.3.2. Recuperación de la solución primal	29
Capítulo 5. Resultados numéricos	31
5.1. Casos de estudio	31
5.1.1. Implementación	33
5.2. Resultados de aplicar el Método del Subgradiente	33
5.3. Resultados de aplicar el Algoritmo de Cutting Plane de Dantzig	34
Capítulo 6. Conclusiones y futuras investigaciones	41
Apéndice A. Valor de los coeficientes $\tilde{\beta}$ y $\tilde{\gamma}$	43
Apéndice B. Matrices origen-destino de Madrid y Barcelona	45
Apéndice. Bibliografía	47

Índice de figuras

1	Algoritmo para purgar el MP	14
2	Representación del grafo $G = (N, A)$.	16
3	Detalle de los nodos y arcos que representan líneas y paradas.	16
4	Configuración tipo de las estaciones	18
5	Representación esquemática de la Red de Madrid	32
6	Representación esquemática de la Red de Barcelona	32
7	<i>relgap</i> método Subgradiente	33
8	Tiempo de ejecución del método de aceleración para el Cutting Plane	35
9	Convergencia del MP y SP	36
10	<i>relgap</i> método Cutting Plane	36

Índice de cuadros

1	Parámetros de las redes estudiadas	32
2	Resultados del Cutting Plane para la red de Madrid	35
3	Resultados del Cutting Plane para la red de Barcelona	35
4	Parámetros líneas seleccionadas Madrid	38
5	Parámetros líneas seleccionadas Barcelona	38
6	Servicios en las paradas para Madrid	39
7	Servicios en las paradas para Barcelona	39
8	Coefficientes $\tilde{\beta}$ y $\tilde{\gamma}$	43
9	Matriz de viajes origen-destino para Madrid	45
10	Matriz de viajes origen-destino para Barcelona	45

Capítulo 1

Introducción

El transporte público es un componente básico en la estructura social, económica y física de un área urbana. Además, se considera un eslabón importante del desarrollo urbano sostenible, ya que debería permitir movimientos más eficientes a lo largo de la ciudad.

Planear, operar y controlar un sistema de transporte público puede resultar desafiante. Están involucrados varios actores con diferentes objetivos, principalmente los usuarios y operadores, donde las necesidades de viaje de los usuarios varían significativamente en espacio y tiempo, de acuerdo a sus características socioeconómicas. Además, la congestión vial y la limitada capacidad vehicular también afectan la decisión y experiencia de cada pasajero [1].

De acuerdo a [2] el problema de Planificación de la Red de Tránsito (TNP, por sus siglas en inglés) abarca las siguientes decisiones que deben tomarse antes de la operación del sistema:

- Planificación estratégica: involucra decisiones a largo plazo, como el diseño de rutas y redes de tránsito
- Planificación táctica: involucra decisiones a mediano plazo, como determinar las frecuencias de servicio a lo largo de las rutas y los horarios.
- Planificación operacional: involucra decisiones que se toman a corto plazo, como la asignación de buses a rutas y de conductores.

El problema del diseño óptimo de rutas y frecuencias para transporte público se denomina TNDP (*Transit Network Design Problem*). En general, resolver el TNDP consiste en encontrar un conjunto de rutas del transporte público, con frecuencias asociadas, de acuerdo a la estructura geográfica y la demanda, con el fin de optimizar los objetivos de los usuarios y operadores, bajo ciertas restricciones que normalmente son: satisfacción de la demanda, niveles de servicio y disponibilidad de recursos.

En este trabajo el TNDP involucra tanto una planificación estratégica como táctica. La decisión estratégica corresponde a seleccionar el conjunto de líneas de autobuses que operará en la extensión del servicio. Las decisiones tácticas están relacionadas

con la mejora del nivel de servicio (minimizar los tiempos de viaje y espera en las paradas de los usuarios) y la reducción de los costos operacionales.

Uno de los elementos importantes considerados para la asignación de pasajeros es el efecto de la capacidad de los vehículos. Cuando una ruta está muy congestionada es necesario dimensionar los servicios, de forma que los pasajeros puedan abordar un autobús y no se vean obligados a dejar pasar un autobús de la línea (o más de uno) por no poder ascender, dado que no tienen suficiente capacidad. Por otro lado, también se considera la capacidad de las paradas para acomodar las colas de pasajeros que esperan abordar los autobuses. En este caso, el efecto de estas limitaciones se incorporan en el modelado a través del impacto del tiempo adicional de espera.

1.1. Revisión literaria

A continuación se presentan los principales modelos de optimización para el TNDP encontrados en la literatura, de acuerdo al estudio realizado por Mautonne en [3]. El autor considera aquellos trabajos que presentan las siguientes características: tienen como variables de decisión los trazados de las rutas y las frecuencias; modelan los intereses de los pasajeros y operadores; y toman en cuenta la demanda, estructura y costos de la red.

Debido a que los modelos para el TNDP mencionados en el estudio de Mautonne presentan una orientación paralela al modelo resuelto en este trabajo de máster, se han reproducido dichos modelos, por completitud, en la siguiente Sección 1.1.2 adoptándose la misma notación que en [3].

1.1.1. Notación

La notación en común utilizada para los trabajos estudiados en [3] es la siguiente. El resto se presentará cuando se requiera. Parte de la notación presentada se ha modificado con el fin de que se asemeje a la utilizada en este trabajo de máster.

- N : conjunto de nodos de la red.
- W : conjunto de pares origen-destino.
- R : conjunto de rutas.
- g_{ij} : demanda entre los nodos en el arco $(i, j) \in W$.
- t_{ij} : tiempo total de viaje en el arco (i, j) , $t_{ij} = t_{ij}^{in} + t_{ij}^w + t_{ij}^t$.
- $t_{ij}^{in}, t_{ij}^w, t_{ij}^t$: son los tiempos de viaje en vehículo, espera y transbordo, en el arco (i, j) , respectivamente.
- f_k : frecuencias de los autobuses operando en la ruta $k \in R$.
- f_{min} : frecuencia mínima de los autobuses operando en cualquier ruta.
- T_k : duración total (ida y vuelta) de la ruta $k \in R$.
- T_{min}, T_{max} : duración mínima y máxima de las rutas, respectivamente.
- r_{max} : máxima cantidad de rutas permitidas.
- n_k : cantidad de autobuses operando en la ruta $k \in R$, $n_k = f_k T_k$.
- p : tamaño de la flota disponible para operar.
- ρ_k : factor de carga de la ruta $k \in R$.

- θ_i : coeficientes de conversión y pesos relativos en la función objetivo.

1.1.2. Modelos

A continuación se presentan algunos de los trabajos relacionados a modelos para el TNDP estudiados en [3], prestando atención a la función objetivo y las restricciones.

1.1.2.1. Modelo de Ceder y Wilson (1986)

En el trabajo de Ceder y Wilson [4], se sugieren dos formulaciones matemáticas diferentes para el TNDP. La primera considera los intereses de los usuarios de minimizar el tiempo de viaje excesivo, expresado como la suma del tiempo de viaje excedente más el tiempo de transbordos para todos los pares origen-destino. La formulación es la siguiente:

$$\begin{aligned} \text{mín} \quad & \left\{ \sum_{i,j \in N} g_{ij}(t_{ij}^{in} - t_{ij}^{min}) + \sum_{i,j \in N} t_{ij}^t \right\} \\ \text{s.a} \quad & \frac{t_{ij}^{in}}{t_{ij}^{min}} \leq 1 + \alpha_{ij}, \quad \forall k \in R \end{aligned} \quad (1)$$

$$T_{min} \leq T_k \leq T_{max}, \quad \forall k \in R \quad (2)$$

$$|R| \leq r_{max} \quad (3)$$

donde t_{ij}^{min} es el tiempo de viaje más corto entre i y j ; y α_{ij} es el desvío máximo permitido para el tiempo de viaje en vehículo de los pasajeros que van de i a j .

La segunda fase presenta un desarrollo más desafiante, dado que considera tanto los intereses de los pasajeros como de los operadores. La formulación propuesta es la siguiente:

$$\begin{aligned} \text{mín} \quad & \left\{ \theta_1 \left(\sum_{i,j \in W} g_{ij}(t_{ij}^{in} - t_{ij}^{min}) + \sum_{i,j \in W} t_{ij}^t + \sum_{i,j \in W} t_{ij}^w \right) + \theta_2 \sum_{k \in R} n_k \right\} \\ \text{s.a} \quad & \text{Rest. (1), (2) y (3)} \\ & f_k \geq f_{min}, \quad \forall k \in R \quad (4) \\ & \sum_{k \in R} n_k \leq p, \quad \forall k \in R \quad (5) \end{aligned}$$

Las restricciones de la primera fase corresponden a la limitación en el tiempo máximo de desplazamiento (como un porcentaje por encima del camino más corto), la limitación de la duración total de la ruta y el número máximo de rutas. Las restricciones que se agregan en la segunda fase tienen relación con la frecuencia mínima y el tamaño máximo de la flota.

1.1.2.2. Modelo de Baaj y Mahmassani (1991)

En [5] los autores proponen minimizar una combinación de los objetivos de los usuarios y operadores. El primero se corresponde con el tiempo total de viaje que incurren los usuarios en la red, mientras que los costos de los operadores están representados por la cantidad de autobuses requeridos para una configuración particular. La formulación del modelo es la siguiente:

$$\begin{aligned} \text{mín} \quad & \left\{ \theta_1 \sum_{i,j \in N} g_{ij} t_{ij} + \theta_2 \sum_{k \in R} n_k \right\} \\ \text{s.a} \quad & \text{Rest. (4), y (5) en el modelo anterior} \\ & \rho_k = \frac{Q_k^{max}}{f_k c} \leq \rho_{max}, \quad \forall k \in R \end{aligned} \quad (6)$$

donde Q_k^{max} es el máximo flujo que ocurre en cualquier arco de la ruta $k \in R$; c es la capacidad de asientos en los autobuses; y ρ_{max} es el máximo factor de carga permitido.

Este modelo utiliza las mismas restricciones de la segunda fase del modelo de Ceder y Wilson y agrega la restricción (6) que establece un máximo valor de ocupación de los autobuses de acuerdo a la capacidad de pasajeros sentados.

1.1.2.3. Modelo de Israeli y Ceder (1993)

La formulación de Ceder y Wilson (1986) fue extendida por Israeli y Ceder en [6], donde el TNDP es formulado como un modelo de programación multiobjetivo, con dos objetivos: el costo total de los pasajeros Z_1 y el tamaño de la flota Z_2 .

$$\begin{aligned} \text{mín} \quad Z_1 &= \left\{ \sum_{i,j \in N} g_{ij} (\theta_1 (t_{ij}^{in} + t_{ij}^t) + \theta_2 t_{ij}^w) + \theta_3 \sum_{k \in R} \gamma_k \right\} \\ \text{mín} \quad Z_2 &= \sum_{k \in R} n_k \end{aligned}$$

donde γ_k es definido como el tiempo en que los asientos de un autobús están vacíos en una ruta $k \in R$. Este término es de interés tanto para los usuarios, por que permite maximizar su comodidad, como para los operadores, dado que tener una alta desocupación de los asientos disponibles no es rentable.

De acuerdo a lo estipulado en [6], en la primera etapa la dimensión del problema se reduce a través de la construcción de un esquema de ruta factible que cumple con una restricción máxima de tiempo de viaje. La segunda etapa se basa en un procedimiento que incorpora procesos de optimización para derivar la función objetivo mínima Z_1 . Mientras se busca minimizar Z_1 , el procedimiento crea varias soluciones Z_2 . Finalmente, el conjunto más deseable de (Z_1, Z_2) se deriva a través de técnicas conocidas en programación multiobjetivo.

1.1.2.4. Modelo de Tom y Mohan (2003)

En [7] los autores presentan prácticamente el mismo modelo que Baaj y Mahmasani (1991). La diferencia es que incorporan las restricciones en la función objetivo y expresan el costo total por unidad de viaje, lo que hace más fácil la comparación entre distintas redes. Por lo que el problema se transforma en la minimización del costo total del sistema por viaje Z , formulado de la siguiente manera:

$$Z = \frac{\theta_1 \left(\sum_{i,j \in N} g_{ij} t_{ij} + \sum \delta_{ij} H \right) + \theta_2 \sum_{k \in R} n_k}{\sum_{i,j \in N} g_{ij}}$$

donde δ_{ij} es la demanda no satisfecha; y H es la duración del período de diseño. El término $\sum \delta_{ij} H$ asigna una penalización igual a la duración del diseño para la demanda insatisfecha. Esto significa que un pasajero que no pudo abordar un autobús experimenta un tiempo de espera igual a la duración del diseño.

1.1.3. Principales características de los modelos para el TNDP

Las principales características de los modelos presentados en [3] para el TNDP tienen relación con que las variables de decisión corresponden a los trazados de las rutas y las frecuencias; en la función objetivo se suele presentar la ponderación económica de los intereses de los pasajeros (minimización de los tiempos de viaje) y operadores (el tamaño de la flota requerida); y las restricciones más comunes son las que acotan las frecuencias, el tamaño de la flota, la duración de las rutas y el factor de carga.

Las diferencias más importantes entre los modelos presentados corresponden a la cantidad de fases, siendo la mayoría en una sola fase; y si es un único objetivo o multiobjetivo, donde se suele resumir los intereses de usuarios y operadores en una sola expresión, mediante los coeficientes de ponderación.

1.2. Objetivos

El principal propósito de este trabajo es presentar una formulación simplificada y métodos de resolución alternativos para el modelo de asignación de frecuencias a líneas auxiliares de transporte público en situaciones de elevada demanda en [8].

Con respecto a los métodos de resolución, el objetivo es estudiar algoritmos de descomposición y proponer métodos que mejoren y/o aceleren la convergencia hacia un resultado óptimo.

Por último, este trabajo tiene como fin validar los métodos de resolución propuestos mediante su aplicación en casos de estudio reales, así como comparar los resultados obtenidos entre los distintos métodos. Para lograr este objetivo, se utilizan dos redes de líneas auxiliares para casos de interrupción del sistema de transporte del corredor ferroviario Paseo de La Castellana (Tunel de la Risa) de Madrid y de la

línea 1 del metro de Barcelona (desde la estación Plaça d'Espanya hasta la estación Clot).

1.3. Organización del documento

El documento se ha organizado de la siguiente forma: en el Capítulo 2 se describen los métodos de descomposición que permitirán resolver el problema. En el Capítulo 3 se describe el problema, la notación utilizada y se explica la formulación matemática proveniente de [8]. El Capítulo 4 explica la aplicación de los métodos presentados en el Capítulo 2 al modelo presentado en el Capítulo 3. En el Capítulo 5 se muestran los resultados numéricos obtenidos al aplicar los métodos de resolución a dos redes de transporte. Finalmente, en el Capítulo 6 se presentan las conclusiones y posibles futuras investigaciones relacionadas.

Capítulo 2

Métodos de descomposición

En este capítulo se estudian los fundamentos y propiedades de la Relajación Lagrangiana, así como los métodos que permiten resolverla, con el objetivo de tener una mejor comprensión de sus características al momento de la implementación en un problema de gran escala.

En primer lugar, se estudian las propiedades de la Programación Lineal y la Relajación Lagrangiana. Luego, se presentan el Método del Subgradiente y el Algoritmo de Cutting Plane, que permitirán resolver la Relajación Lagrangiana.

2.1. Programación Lineal

Un *Problema Lineal* (LP, por sus siglas en inglés) consiste en encontrar un $x \in \mathbb{R}^n$ que satisfaga un conjunto de m restricciones lineales. Por lo general se expresa así

$$\min_x f(x) \tag{7}$$

$$s.a \quad Ax \leq b \tag{8}$$

$$x \in \mathbb{R}^n \tag{9}$$

donde A es una matriz $(m \times n)$ sobre \mathbb{R} , y $b \in \mathbb{R}^m$.

Cada $x \in \mathbb{R}^n$ es denominado una *solución* del LP. Si x satisface las restricciones (8) es una *solución factible*. El *valor objetivo* del LP para un dado x es $f(x)$. Entre todas las soluciones factibles existe una *solución óptima*¹ x^* que produce el mínimo, es decir, el óptimo valor objetivo $f(x^*)$.

Si se reemplazan las restricciones (9) con las restricciones de integridad $x \in \mathbb{N}^n$, el problema es llamado un *Problema de Programación Entera* (ILP, por sus siglas en inglés). Si algunas de las variables tienen un valor real y algunas son enteras sería un *Problema Entero Mixto* (MIP, por sus siglas en inglés). Una especial variante de

¹La solución óptima no necesariamente es única.

un ILP es un 0-1 ILP (o binario), el cual es definido de la siguiente manera:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.a.} \quad & Ax \leq b \\ & x \in \{0, 1\}^n \end{aligned}$$

Los LP pueden ser resueltos eficientemente, por ejemplo, por el método simplex o el método del punto interior. Los ILP y MIP son más difíciles de resolver. Los métodos generales para resolver de manera óptima los ILP y MIP son las aproximaciones *Branch and Bound* o *Branch and Cut* (ver [9]). En entornos de gran escala una buena forma de obtener cotas inferiores es mediante la relajación Lagrangiana, la cual se describirá en la siguiente sección.

2.2. Relajación Lagrangiana

La *Relajación Lagrangiana* (LR, por sus siglas en inglés) es un método que calcula cotas inferiores para los LP², ILP o MIP. El fundamento teórico del método se puede encontrar en [10], [11], [12], entre otros.

La idea de un LR es relajar algunas de las restricciones y “moverlas” a la función objetivo. Sea (P) un problema general de optimización definido como

$$(P) \quad \min_x \quad f(x) \tag{10}$$

$$\text{s.a.} \quad Ax \leq b \tag{11}$$

$$Bx \leq d \tag{12}$$

$$x \in X \tag{13}$$

donde (11) contiene las restricciones que se consideran complicadas, de modo que sin ellas el problema sería mucho más fácil de resolver. Las restricciones (12) y (13) se mantendrán en el problema. La expresión (13) contiene las restricciones de signo e integridad.

El LR resultante de la relajación de las restricciones (11) será

$$(LR_\mu) \quad \min_x \quad f(x) + \mu(Ax - b) \tag{14}$$

$$\text{s.a.} \quad Bx \leq d \tag{15}$$

$$x \in X \tag{16}$$

Aquí $\mu \in \mathbb{R}_{\geq 0}^{m_A}$ es cualquier vector positivo. Sus componentes son denominados *multiplicadores de Lagrange*. Intuitivamente, se puede ver que μ puede imponer una penalización a las restricciones violadas. El valor objetivo óptimo de (LR_μ) es una cota inferior para el valor objetivo óptimo de (P). Beasley demuestra este hecho en [13] de la siguiente manera.

²Aunque en general los LP pueden resolverse eficientemente, el LR puede ser interesante en ciertos LP que son extremadamente grandes o, de otro modo, demasiado complejos para ser resueltos directamente.

El valor objetivo óptimo de (P) no es menor que el valor objetivo óptimo de (P')

$$(P') \quad \min_x \quad f(x) + \mu(Ax - b) \quad (17)$$

$$s.a \quad Ax \leq b \quad (18)$$

$$Bx \leq d \quad (19)$$

$$x \in X \quad (20)$$

dado que $\mu(Ax - b) \leq 0$. Esto a su vez no es más pequeño que el valor objetivo óptimo del problema (LR_μ) , ya que eliminar restricciones en un problema de minimización solo puede conducir a un valor objetivo óptimo más pequeño. En simples palabras

$$v(LR_\mu) \leq v(P') \leq v(P) \quad (21)$$

El siguiente paso es encontrar el vector μ^* , que produce el límite inferior más grande, es decir, el límite inferior que es lo más cercano posible al valor de la solución óptima. Este es denominado problema *Dual Lagrangiano* (DL).

$$(DL) \quad \max_{\mu \geq 0} \left\{ \begin{array}{ll} \min_x & f(x) + \mu(Ax - b) \\ s.a & Bx \leq d \\ & x \in X \end{array} \right\}$$

Existen algoritmos heurísticos para encontrar buenos valores para los multiplicadores de Lagrange, tales como el *Método del Subgradiente* y el *algoritmo de Cutting Plane* descritos en la siguientes secciones.

2.3. Método del Subgradiente

En esta subsección, se describirá un método para aproximar el DL

$$\mathcal{L}^* = \max_{\mu} \mathcal{L}(\mu) \quad (22)$$

de la función Lagrangiana que relaja las restricciones $Ax \leq b$, que es

$$\mathcal{L}(\mu) = \min \left\{ f(x) + \mu(Ax - b) \mid x \in X \right\} \quad (23)$$

El principio de la relajación Lagrangiana es incluir las restricciones “difíciles” en la función objetivo. En otras palabras, se supone que optimizar una función sobre X es “fácil”. Por lo tanto, si μ es fijo, se puede calcular el valor de $\mathcal{L}(\mu)$ y una solución óptima correspondiente a $x \in X$.

Hay que tener en cuenta que la función Lagrangiana es la envolvente inferior del conjunto de hiperplanos $\{f(x) + \mu(Ax - b) \mid x \in X\}$. Por lo tanto, $\mathcal{L}(\mu)$ es una función cóncava, lo que es equivalente a resolver el siguiente problema lineal:

$$\begin{array}{ll} \max & z \\ s.a & z \leq f(x) + \mu(Ax - b) \\ & x \in X, \mu \in \mathbb{R}^n \end{array} \quad (24)$$

Generalmente el número de restricciones de dicho programa es exponencial, por lo que se utiliza un método de descenso del gradiente para calcular un valor lo mas cercano posible al dual Lagrangiano. Dada una función cóncava $f : \mathbb{R}^n \rightarrow \mathbb{R}$, un vector g es un *subgradiente* de f en x si para cualquier $y \in \mathbb{R}^n$, $f(y) \leq f(x) + g^T(y - x)$. La función f es *diferenciable* en x si y solo si f admite un único subgradiente en x . Si \mathcal{L} fuera diferenciable, se usaría el método de descenso del gradiente para converger hacia el valor óptimo. Sin embargo, en este caso, \mathcal{L} no es diferenciable en todas partes, porque es una función lineal por partes. Por lo que se calcula una secuencia de $(\mu^{(k)})_{k \in \mathbb{N}}$ tal que $\mathcal{L}(\mu^{(k)})$ converge hacia la solución óptima, usando el siguiente algoritmo.

Algoritmo 1 Método del Subgradiente

- 1: Hacer $k = 0$ y elegir $\mu^{(0)} \in \mathbb{R}^n$
 - 2: Calcular $\mathcal{L}(\mu^{(k)})$ y $x^{(k)} \in X$
 - 3: Elegir un subgradiente $g^{(k)} = Ax^{(k)} - b$ de la función \mathcal{L} en $\mu^{(k)}$
 - 4: **Si** $g^{(k)} = 0$ **entonces** STOP, la solución óptima es $\mathcal{L}(\mu^{(k)})$
 - 5: Calcular la longitud de paso $\alpha^{(k)}$
 - 6: Calcular $\mu^{(k+1)} = \mu^{(k)} + \alpha^{(k)} g^{(k)}$
 - 7: Hacer $k = k + 1$ y volver al paso 2.
-

Cuando las restricciones del problema inicial que se relajan son de la forma $Ax \leq b$, en el paso 6 μ corresponde al máximo entre 0 y $\mu + \alpha g$, dado que $\mu \geq 0$.

2.3.1. Longitud de paso

Usualmente la fórmula para calcular la longitud de paso es

$$\alpha^{(k)} = \frac{UB - \mathcal{L}(\mu^{(k)})}{\|Ax^{(k)} - b\|^2}, \quad k = 1, 2, \dots \quad (25)$$

donde UB es una cota superior de la solución óptima que se quiere calcular, que puede ser obtenida aplicando una heurística al problema primal. Sin embargo, a veces es difícil encontrar buenas cotas superiores, que garanticen la convergencia del método, por lo que existen otras formas de calcular la longitud de paso.

En [14] se desarrollan diferentes estrategias para elegir la longitud de paso, que pueden ser utilizadas en este contexto, tales como la utilizada en el *Método de medias ponderadas sucesivas* (MSWA, por sus siglas en inglés) y el *Método de media autorregulada*.

El más simple es la longitud de paso para el MSWA y es de la forma

$$\alpha^{(k)} = \frac{k^d}{\sum_{i=1}^k i^d}, \quad \forall k = 1, 2, \dots \quad (26)$$

donde el parámetro de ponderación $d \geq 0$ es un número real.

En el método de media autorregulada las longitudes de paso se actualizan evaluando una función potencial. Cuando la función potencial detecta que la longitud de paso en la iteración anterior es efectiva para la convergencia, mantiene la longitud de

paso actual convergiendo lentamente a cero, de lo contrario, acelera la reducción de la longitud de paso.

Para el método del subgradiente, la medida más conveniente que se puede usar para monitorear la convergencia es mediante la norma del subgradiente, de la siguiente manera

$$\beta^{(k)} = \begin{cases} \beta^{(k-1)} + \Gamma, & \Gamma > 1, \text{ si } \|g^{(k)}\| \geq \|g^{(k-1)}\| \\ \beta^{(k-1)} + \gamma, & \gamma < 1, \text{ si } \|g^{(k)}\| < \|g^{(k-1)}\| \end{cases} \quad (27)$$

siendo la longitud de paso

$$\alpha^{(k)} = \frac{1}{\beta^{(k)}}, \quad \forall k = 1, 2, \dots \quad (28)$$

La elección de los parámetros de incremento de la longitud de paso Γ y γ es flexible, por ejemplo, $\Gamma \in [1.5, 2]$ y $\gamma \in [0.01, 0.5]$.

El método de promediado autorregulado se basa en la consideración de que la longitud de paso debe ser mayor para permitir una exploración más agresiva del espacio de la solución cuando la iteración actual converja, de lo contrario, debería ser menor cuando las soluciones divergen. La serie de longitud de paso $\{\alpha^{(k)}\}$ en el método de promediado autorregulado satisface las condiciones de $\sum_{k \geq 0} \alpha^{(k)} = \infty$ y $\lim_{k \rightarrow \infty} \alpha^{(k)} = 0$, que garantizan la convergencia del método.

2.3.2. Criterio de parada

Cuando se resuelven problemas de gran escala suele ser difícil obtener $g^{(k)} = 0$ en el algoritmo [1](#). Una buena alternativa para que el proceso iterativo se detenga, es el gap relativo *relgap*, que se calcula de la siguiente forma

$$relgap^{(k)} = \frac{gap^{(k)}}{\|\mu^{(k)}\| \|g^{(k)}\| + \epsilon}, \quad k = 1, 2, \dots \quad (29)$$

donde $gap^{(k)} = \mu^{(k)} g^{(k)}$, $k = 1, 2, \dots$, y ϵ es un valor cercano a cero (por ejemplo 10^{-8}) que evita un error en la ecuación cuando la norma de los vectores es cero.

De esta forma, el algoritmo se detendrá cuando $relgap \leq \varepsilon$, siendo ε , por ejemplo, 10^{-2} .

Adicionalmente, se puede establecer como criterio de parada si un número máximo de iteraciones K y/o el tiempo máximo de ejecución es alcanzado.

2.3.3. Recuperación de la solución primal

Si bien el método del subgradiente puede ser bastante poderoso para proporcionar de forma rápida un límite inferior para el LP mediante el DL, la desventaja es que una solución primal óptima (o incluso una solución factible y casi óptima) para el LP no suele obtenerse a través de este proceso. Para resolver esta problemática, en [15](#) se plantea una aproximación que pondera las soluciones y las longitudes de paso obtenidas en cada iteración para recuperar la solución primal del problema

original. El promedio ponderado es de la forma

$$\bar{x}^{(k)} = \sum_{i=1}^k \bar{\alpha}^{(i)} x_i, \quad \text{donde } \bar{\alpha}^{(i)} = \frac{\alpha^{(i)}}{\sum_{j=1}^k \alpha^{(j)}}, \quad 1 \leq i \leq k. \quad (30)$$

Se puede observar que para cada $k \geq 1$ se tiene $\bar{\alpha}^{(i)} \geq 0$, $i = 0, \dots, k$ y $\sum_{i=1}^k \bar{\alpha}^{(i)} = 1$.

2.4. Algoritmo de Cutting Plane de Dantzig

El algoritmo de Cutting Plane de Dantzig (1959) consiste en descomponer el problema original en un “*problema maestro*” (MP, por sus siglas en inglés) y un “*subproblema*” (SP), de modo que la solución del MP forma parte de la definición del SP, y la solución del SP genera planos de corte, o simplemente cortes, para ir acotando la solución del MP.

En simples palabras, el SP resuelve la relajación Lagrangiana del problema original y la solución obtenida se añade al MP para actualizar los multiplicadores de Lagrange. Este proceso se repite hasta encontrar una solución óptima. El algoritmo se presenta a continuación.

Algoritmo 2 Cutting Plane de Dantzig

- 1: Hacer $k = 0$. Encontrar un punto $x^{(0)} \in X$
- 2: Resolver el MP:

$$\begin{aligned} \max_{z, \mu} \quad & z \\ \text{s.a.} \quad & z \leq f(x^{(i)}) + \mu(Ax^{(i)} - b) \quad i = 0, \dots, k \\ & \mu \geq 0, z \text{ irrestricto} \end{aligned}$$

y obtener $(z^{(k)}, \mu^{(k)})$

- 3: Resolver el SP:

$$\begin{aligned} w(\mu^{(k)}) &= \min_{x \in X} \{f(x) + \mu^{(k)}(Ax - b)\} \\ &= f(x^{(k+1)}) + \mu^{(k)}(Ax^{(k+1)} - b) \end{aligned}$$

y obtener $(x^{(k+1)})$

- 4: **Si** $z^{(k)} \leq w(\mu^{(k)})$ **entonces** STOP, $\mu^{(k)}$ es el óptimo del problema dual que satisface todas las restricciones del MP.
- 5: **Si** $z^{(k)} > w(\mu^{(k)})$ **entonces** agregar al MP la restricción

$$z \leq f(x^{(k+1)}) + \mu^{(k)}(Ax^{(k+1)} - b)$$

y hacer $k = k + 1$ e ir al paso 2

2.4.1. Criterio de parada

Cuando se trabaja con problemas de gran escala, puede resultar difícil lograr que

$z^{(k)} \leq w(\mu^{(k)})$, por lo que existe un método alternativo para detener el proceso iterativo mediante el cálculo del gap relativo:

$$relgap = \frac{z^{(k)} - w(\mu^{(k)})}{w(\mu^{(k)})} \leq \varepsilon_{mp} \quad (31)$$

donde ε_{mp} es la tolerancia que se le exige al método.

2.4.2. Obtención de una solución primal factible

Cabe señalar que, luego de finalizar, el algoritmo de Dantzig puede que no genere una solución factible primal para el problema original. Para resolver esta situación, se genera una solución factible próxima a la solución óptima a partir de las soluciones obtenidas en cada iteración. El enunciado formal y la demostración pueden verse en [16].

Se plantea el siguiente problema de optimización

$$\begin{aligned} \bar{z}^{(k)} = \quad & \min_{\alpha} \quad \sum_{i=0}^k \alpha^{(i)} f(\hat{x}^{(i)}) \\ \text{s.a} \quad & \sum_{i=0}^k \alpha^{(i)} g(\hat{x}^{(i)}) \leq 0 \\ & \sum_{i=0}^k \alpha^{(i)} = 1, \alpha^{(i)} \geq 0 \end{aligned} \quad (32)$$

donde $g(\hat{x}^{(i)}) = A\hat{x}^{(i)} - b$ y $\hat{x}^{(i)}, i = 0, \dots, k$, son las soluciones que se obtuvieron con el algoritmo de Dantzig.

Sea $\alpha^{*(i)}, i = 0, \dots, k$, las soluciones al problema anterior y si $\tilde{x} = \sum_{i=0}^k \alpha^{*(i)} \hat{x}^{(i)}$, se verifica que $\bar{z}^{(k)} - w(\mu^{(k)}) \leq \epsilon \Rightarrow f(\tilde{x}) \leq z^* + \epsilon$, donde z^* es el valor óptimo de la función objetivo

$$\begin{aligned} z^* = \min_{x \in X} \quad & f(x) \\ & g(x) \leq 0 \end{aligned}$$

2.4.3. Aceleración del algoritmo de Cutting Plane

El MP en el algoritmo [2] tiene un infinito número de restricciones, correspondientes a los cortes, del tipo $z \leq f(x^{(i)}) + \mu(Ax^{(i)} - b)$, $i = 0, \dots, k$. Para evitar esto, existe un algoritmo que permite purgar el MP y acelerarlo. De acuerdo con los estudios de Lawphongpanich y Hearn (ver [17]) cuando el SP es fácil de resolver, es ventajoso resolver más subproblemas con el fin de obtener mejores cortes, es decir, aquellos que pueden conducir a una reducción en el número de MP a resolver. El algoritmo de purgado se describe en la Figura [1].

Otro método para acelerar el algoritmo, es actualizando los multiplicadores de Lagrange de la siguiente forma:

$$\mu^{(k+1)} = \mu^{(k)} + \alpha^{(k)} \mu^*$$

donde μ^* es la solución obtenida al resolver el MP, y la longitud de paso α se obtiene mediante el método MSWA, explicado en la sección 2.3.1. Sin embargo, es necesario reiniciar el valor de la longitud de paso cada cierto número de iteraciones, por ejemplo cada 10, con el fin de evitar el estancamiento de la iteración anterior.

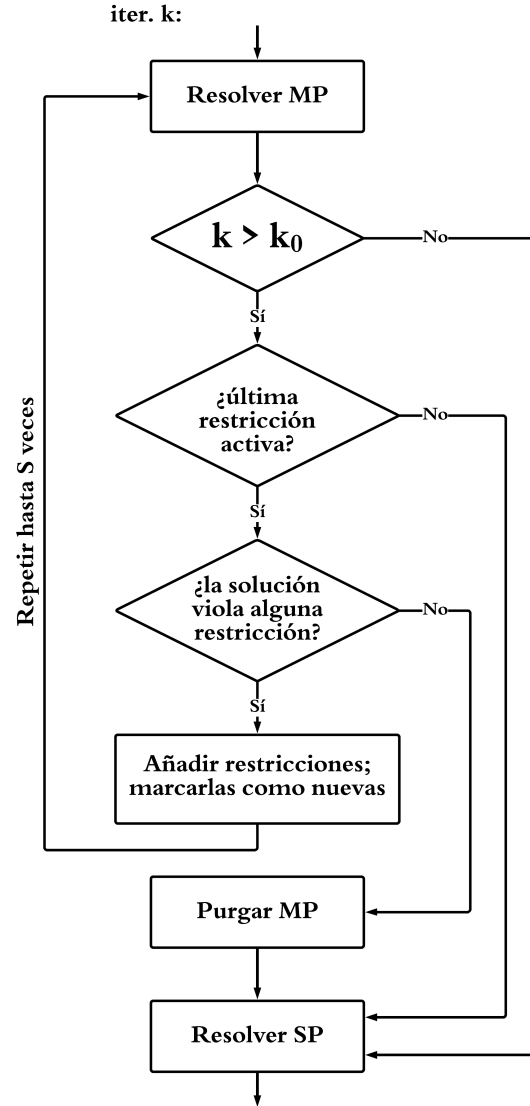


Figura 1. Algoritmo para purgar el MP

Capítulo 3

Definición del problema

En este capítulo se describe un modelo basado en programación matemática para ayudar con la planificación eficiente de servicios para un conjunto de líneas de autobuses auxiliares. Este modelo corresponde a una simplificación del presentado en [8], el cual es capaz de cubrir casos de elevada congestión, mientras que el que se presenta aquí es para redes con congestión mediana. Por comodidad, se describe en detalle el modelo en [8] y las simplificaciones efectuadas en él.

La información de entrada básica que necesita el modelo es (a) un conjunto de líneas de autobuses, en un entorno urbano, identificado por sus segmentos y paradas, que sean candidatas para operar y (b) una tabla de viajes origen-destino para las paradas afectadas.

El modelo minimiza los costos de implementación y operación del servicio más el valor económico del tiempo total de viaje de los pasajeros.

Las soluciones que entrega el modelo son (a) el número de autobuses a asignar a cada línea candidata, (b) las rutas que debe seguir el pasajero entre cada par origen-destino, y (c) el tiempo de espera de los pasajeros en las paradas.

3.1. Modelo de red

Los flujos de pasajeros transcurrirán sobre un grafo dirigido $G = (N, A)$ cuya estructura aparece esbozada en las Figura 2 y Figura 3. El conjunto de nodos N se divide en dos subconjuntos: N_G que son los nodos que aparecen en el plano y por otra parte esta $N \setminus N_G$ que son todos los demás y que conforman las líneas de autobuses. Los nodos en N_G y los arcos $a = (i, j)$ tales que $i \in N_G, j \in N_G$ se usan para representar la red de transbordos.

Por \hat{N}_G se designan los nodos en N_G que actúan como paradas. L es el conjunto de líneas. Por Π_ℓ se designa el conjunto de n_ℓ paradas que definen una línea $\ell \in L$, es decir, $\Pi_\ell = \{b_1, \dots, b_{n_\ell}\}$. Por L_b se designa el conjunto de líneas que contienen la parada $b \in \hat{N}_G$. Se define el conjunto A_b como los arcos cuyos nodos no están en N_G y que acceden a la parada b :

$$A_b = \{(i, j) \in A \mid i, j \notin N_G, \exists (j, k) \in E(j) \text{ t.q. } \ell \in I(b)\} \quad (33)$$

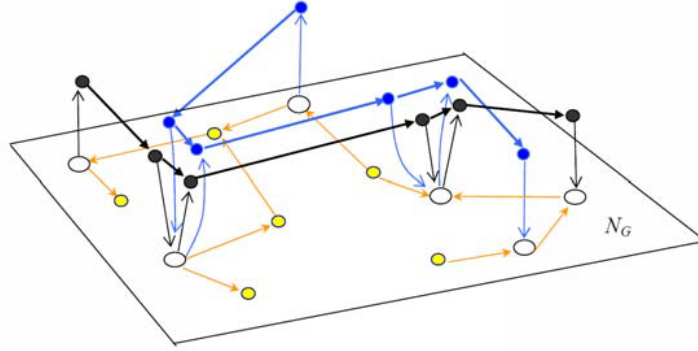


Figura 2. Representación del grafo $G = (N, A)$.

Por A_ℓ se designa el conjunto de arcos cuyos nodos no están en N_G y que modelizan la línea $\ell \in L$:

$$A_\ell = \{(i, j) \in A \mid i, j \notin N_G, (i, j) \text{ pertenecen a la línea } \ell\} \quad (34)$$

$A_{b,\ell}$ define el conjunto de arcos $a \in A_b$ correspondientes a la línea $\ell \in L$.

$$A_{b,\ell} = A_b \cap A_\ell \quad (35)$$

Por A_G se designa el conjunto de arcos en “tierra”:

$$A_G = \{(i, j) \in A \mid i \in N_G, j \in N_G\} \quad (36)$$

En la Figura 3 se muestra en detalle la modelización de las líneas y paradas. Los arcos $a = (b, j)$, $b \in \Pi_\ell$, $j \notin N_G$ son para modelizar el tiempo de abordaje y el de espera en la parada por parte de los pasajeros que abordan la línea ℓ . Dada una parada b en la que se detenga la línea ℓ , el arco de abordaje desde $b \in \hat{N}_G$ hasta la línea ℓ vendrá designado por $a(\ell, b)$. Para un arco de abordaje a , se designa el arco $x(a)$ para los pasajeros que se mantienen en el autobús al estacionarse en la parada y por $y(a)$ se designa el correspondiente arco de bajada.

El conjunto de líneas que inciden en una parada b puede definirse como:

$$L_b = \{\ell \in L \mid A_{b,\ell} \neq \emptyset\} \quad (37)$$

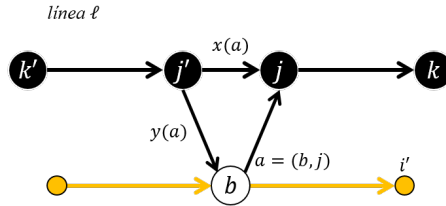


Figura 3. Detalle de los nodos y arcos que representan líneas y paradas.
 $b, i' \in N_G$. $(k', j') \in A_b$. $x(a) = (j', j)$, $(j, k) \in A_\ell$.

3.2. Flujo de pasajeros

La demanda de pasajeros está representada por el conjunto de flujos positivos, $g_\omega > 0$, para cada par origen-destino $\omega = (p, q)$, durante un predefinido periodo de tiempo de largo H unidades. El conjunto de pares origen-destino W está definido a partir de nodos dentro de N_G que son paradas. D es el conjunto de nodos que son destinos. La matriz de viajes origen-destino será designada por $g_\omega, \omega \in W$.

El flujo de pasajeros durante un periodo de tiempo específico será organizado en commodities, uno por cada par origen-destino, y el flujo total de pasajeros en el arco $a \in A$, durante ese periodo de tiempo y para un par origen-destino $\omega \in W$ será designado por v_a^ω . La ecuación de balance de flujos por par origen-destino $\omega = (p, q) \in W$ en un nodo $i \in N$ será la siguiente:

$$\sum_{j \in E(i)} v_{ij}^\omega - \sum_{k \in I(i)} v_{ki}^\omega = \begin{cases} -g_{(p,q)} & i = q \\ g_{(p,q)} & i = p \\ 0 & i \neq p \end{cases} \quad (38)$$

Agregando las condiciones de no-negatividad $v_a^\omega \geq 0$ para los flujos en los arcos de las relaciones previas (38), el siguiente poliedro es definido:

$$V^\omega = \left\{ v^\omega \in \mathbb{R}_+^{|A|} \mid v^\omega = (\dots, v_{ij}^\omega, \dots; (i, j) \in A), v_{ij}^\omega \text{ verifica (38)} \right\}, \quad \omega \in W \quad (39)$$

Finalmente, el poliedro \mathcal{V} del flujo total de pasajeros en los arcos será

$$\mathcal{V} = \left\{ v \in \mathbb{R}_+^{|A|} \mid v = \sum_{\omega \in W} v^\omega, v^\omega \in V^\omega \right\} \quad (40)$$

3.3. Modelado de los tiempos operacionales

Se asumirá que hay p unidades de autobuses disponibles a ser asignados a $|L|$ predefinidas líneas candidatas, con todas las unidades de igual tipo y con capacidad de c pasajeros. Así, si $n^\ell \geq 0, n^\ell \in \mathbb{Z}, \ell \in L$ es el número de unidades asignadas a la línea ℓ , entonces

$$\sum_{\ell \in L} n^\ell \leq p \quad (41)$$

3.3.1. Modelado de las paradas de autobús

La operación de los autobuses en las paradas se modelará usando dos sistemas de espera en tándem con bloqueo como se muestra en la Figura 4. Los sistemas de espera de entrada y salida serán designados como sistema L0 y sistema L1, respectivamente. En este trabajo, se asumirá que el tiempo medio w_b^1 que un autobús está en el espacio L1 de la parada $b \in \hat{N}_G$ es fijo, el cual incluye el tiempo de despejar el puesto de servicio y el tiempo de reingreso al tráfico externo. Además, se asumirá que no existe congestión, por lo que no hay cola de espera en el espacio L0, por lo tanto, $w_b^0 = 0$.

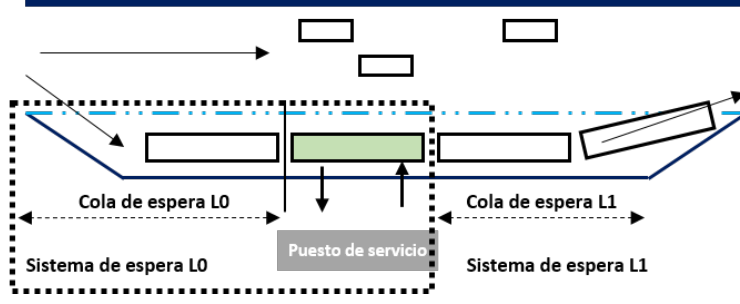


Figura 4. Configuración tipo de las estaciones

3.3.1.1. Tiempos de servicio de autobús y de pasajeros en las paradas

Se asumirá que el tiempo promedio de servicio efectivo κ_b en la parada $b \in \hat{N}_G$ es conocido y se calcula de la siguiente forma:

$$\kappa_b = \frac{\kappa_b^M + \kappa_b^{OM} + \kappa_b^{PST}}{e(s_b)}, \quad b \in \hat{N}_G \quad (42)$$

donde κ_b^M es el tiempo de maniobra del autobús para entrar y salir del puesto de servicio, κ_b^{OM} es un margen operacional y κ_b^{PST} es el tiempo de servicio de pasajeros requerido para abordar/descender del autobús, incluyendo el tiempo de abrir/cerrar puertas. En casos de paradas lineales en las que el largo de los puestos de servicio permite más de un área de carga donde dos o más autobuses pueden ser servidos simultáneamente, un factor de eficiencia e para el puesto debe ser utilizado, el cual depende del máximo número de áreas de carga s_b que es posible asignar.

3.3.1.2. Ciclos de línea y cola de autobuses en las paradas

Asumiendo que n^ℓ autobuses han sido asignados a la línea ℓ del total de p autobuses disponibles y que ellos realizan un total de z^ℓ servicios en la línea $\ell \in L$, el factor de carga ρ_b en la parada b para el sistema L0 de espera será expresado por la relación entre las llegadas de autobuses por unidad de tiempo y la capacidad máxima efectiva del puesto de servicio o la cantidad máxima de servicios del puesto por unidad de tiempo:

$$\rho_b = \frac{\text{Llegada de autobuses por unidad de tiempo}}{\text{Máximo número de servicios por unidad de tiempo}}, \quad b \in \hat{N}_G \quad (43)$$

La llegada de autobuses por unidad de tiempo a la parada b durante un periodo de H unidades de tiempo debe ser $\frac{1}{H} \sum_{\ell \in L_b} z^\ell$; sin embargo, el máximo número de servicios por unidad de tiempo será s_b/κ_b . El factor de carga será entonces expresado como:

$$\rho_b = \frac{\kappa_b}{s_b H} \sum_{\ell \in L_b} z^\ell, \quad b \in \hat{N}_G \quad (44)$$

Los ciclos de autobús C_ℓ en la línea $\ell \in L$ se obtienen de:

$$C_\ell = C_\ell^0 + \sum_{b \in \Pi_\ell} (\kappa_b + w_b^1 + w_b^0), \quad \ell \in L \quad (45)$$

donde C^0 es el resultado del tiempo de viaje libre entre parada y parada. Este tiempo de viaje libre será representado por t_a^0 en el arco $a \in A_\ell^0$, el cual es fijado a priori. Este arco $a = (j_\ell(b), j'_\ell(b'))$ va desde la parada b hasta la parada b' . Por lo tanto, t_a^0 es el tiempo transcurrido comenzando inmediatamente después de dejar la parada b y hasta llegar a la parada b' , entonces,

$$C_\ell^0 = \sum_{a \in A_\ell^0} t_a^0, \quad \ell \in L \quad (46)$$

Otro supuesto es que los tiempos muertos de los autobuses por cambiar de conductor, etc. durante el horizonte temporal H son realizados en depósitos ubicados fuera de cualquier parada en la red y que los autobuses son retirados de la línea durante esos tiempos muertos. Esto hace que el tiempo operacional efectivo para los autobuses asignados a la línea $\ell \in L$, H^ℓ sea menor o igual que H . La relación entre el número de servicios z^ℓ y el número de autobuses n^ℓ asignados a la línea $\ell \in L$ es

$$n^\ell H^\ell \geq z^\ell C_\ell, \quad \ell \in L \quad (47)$$

3.3.1.3. Máximo flujo de entrada de autobuses en una parada

El máximo flujo de entrada de autobuses que una parada puede ser capaz de servir es determinado preliminarmente por dos factores. El primer factor es el tiempo de servicio del autobús, el cual depende del flujo de pasajeros que llega a la parada y el tiempo requerido por los autobuses para acercarse al puesto de servicio más el tiempo de despeje. El segundo factor es la limitación física del espacio en el cual los autobuses pueden hacer cola para alcanzar el puesto de servicio de la parada. Sin embargo, como en este caso no se considera congestión, se asumirá que no hay cola de espera en el sistema L0.

El modelo limita el máximo flujo de entrada de autobuses que puedan entrar en la parada. Se utilizará una estimación obtenida por simulación de la relación entre el promedio del largo de la cola y el máximo largo de la cola que es solo superado un 5 % de las veces. Con ese fin, $\mathcal{L}^0, \mathcal{L}^1, \eta^0$ y η^1 se definen como:

- \mathcal{L}^0 : Máximo número de autobuses permitidos en el espacio L0 (es decir, aquellos permitidos en el puesto de servicio).
- \mathcal{L}^1 : Máximo número de autobuses permitidos en el espacio L1 (es decir, aquellos que están a la salida del puesto de servicio y que esperan el reingreso al tráfico).
- η^0 y η^1 : Factor de ocupación al 95 % o relación entre el promedio de ocupación de los espacios L0 y L1 y la máxima ocupación alcanzada el 95 % de las veces (usualmente entre 3 y 4). Un valor de 3 es adoptado.

Aplicando la fórmula de Little a ambos sistemas de espera L0 y L1 en la parada $b \in \hat{N}_G$:

$$\frac{(\kappa_b + w_b^0)}{H} \sum_{\ell \in L_b} z^\ell \leq \frac{\mathcal{L}^0}{\eta^0}, \quad b \in \hat{N}_G \quad (48)$$

$$\frac{w_b^1}{H} \sum_{\ell \in L_b} z^\ell \leq \frac{\mathcal{L}^1}{\eta^1}, \quad b \in \hat{N}_G \quad (49)$$

Además, si ρ_b es el factor de carga por parada, es deseable que $\rho_b < 1$. Tomando, por ejemplo, $\epsilon = 0.02$ o 0.05 :

$$\frac{\kappa_b}{s_b H} \sum_{\ell \in L_b} z^\ell = \rho_b \leq 1 - \epsilon, \quad b \in \hat{N}_G \quad (50)$$

En resumen,

$$\sum_{\ell \in L_b} z^\ell \leq \hat{Z}_b \triangleq H \min \left\{ \frac{(1 - \epsilon)s_b}{\kappa_b}, \frac{\mathcal{L}^0}{\eta^0(\kappa_b + w_b^0)}, \frac{\mathcal{L}^1}{\eta^1 w_b^1} \right\}, \quad b \in \hat{N}_G \quad (51)$$

3.3.1.4. Limitaciones en el máximo headway y frecuencias admisibles en las líneas

Se requiere que las soluciones de este modelo proporcionen una serie de servicios z^ℓ en la línea $\ell \in L$ de modo que las frecuencias en esa línea caigan dentro de un rango predeterminado de frecuencias, digamos $[\underline{f}, \hat{f}]$, siempre que $z^\ell > 0$. La cota inferior \underline{f} para las frecuencias se expresará en términos del máximo headway h_{max} , es decir, $\underline{f} = H/h_{max}$. Las siguientes restricciones harían cumplir estas condiciones:

$$z^\ell \geq \lambda^\ell \frac{H}{h_{max}}, \quad z^\ell \leq \lambda^\ell H \hat{f} \quad \lambda^\ell \in \{0, 1\}, \quad \ell \in L \quad (52)$$

Claramente, si $\lambda^\ell = 1$, se asignará un número positivo de servicios a la línea ℓ , el headway resultante será $H/z^\ell \geq h_{max}$, y la frecuencia en esta línea no excederá \hat{f} , mientras que si $\lambda^\ell = 0$, no se asignarán servicios a la línea.

3.3.2. Tiempo total de espera de los pasajeros en la parada

El tiempo de espera de los pasajeros en la parada para abordar un autobús puede no ser un componente insignificante en el tiempo total de viaje. Es comúnmente aceptado que los pasajeros que esperan en una parada de autobús puedan ser modelados con un proceso de colas siguiendo las siguientes características:

- (1) Los pasajeros que esperan se encuentran en una cola de servicio en lotes, es decir, las llegadas de los pasajeros son uno a uno, pero hay un servicio por lotes para cada llegada de autobús a la parada. Por lo general, el tiempo total utilizado para el servicio de autobús es más corto que el tiempo medio entre llegadas de autobuses en la parada.
- (2) El número de pasajeros que se puede asignar en cada autobús que llega es una variable aleatoria que es independiente del número de pasajeros que esperan al momento de la llegada.

La llegada de los pasajeros a las paradas de autobús se ha adoptado como poissoniana. En el contexto de la teoría de colas el modelo más adecuado parece ser el

$M/G^{[Y]}/s$. Mediante simulación se han examinado diversas distribuciones de probabilidad para los tiempos entre llegadas de servidores. En general, se ha encontrado que la respuesta es parecida a la que proporciona el clásico modelo $M/M^{[Y]}/1$, es decir, llegadas de pasajeros poissonianas con tasa λ pasajeros por unidad de tiempo, llegadas de autobuses poissonianas con tasa μ y un solo puesto de servicio para los autobuses, $s = 1$. En este caso, el tiempo promedio de espera por pasajero φ para una línea de autobús en una parada viene dado como una función del factor de carga ϱ de la cola de la forma:

$$\varphi(\varrho) = P_0 \xi(\varrho) \quad (53)$$

donde P_0 es el tiempo de espera en la parada de autobús por servicio y pasajero en una situación sin congestión y $\xi(\varrho)$ es una función creciente y convexa del factor de carga ϱ en la parada o el cociente entre la tasa de llegadas de pasajeros y la tasa de pasajeros que pueden aligerar los servidores que llegan a la parada. P_0 depende de la media y desviación del tiempo entre llegadas de autobuses, h y σ , respectivamente:

$$P_0 = \frac{h}{2} \left(1 + \frac{\sigma^2}{h^2} \right) \quad (54)$$

De acuerdo a las investigaciones realizadas en [8], el coeficiente de variación $\chi = \sigma/h$ se asumirá fijo a priori en el intervalo $[0.40, 0.74]$.

Para una línea $\ell \in L$ y una parada de autobús $b \in \Pi_\ell$, sea $a = a(\ell, b)$ el correspondiente arco de la red expandida. El tiempo promedio entre llegadas será H/z^ℓ . El factor de carga ϱ_a para la cola de pasajeros dispuestos a abordar servicios de la línea ℓ estará dado como una función de los flujos en los arcos de la red expandida por

$$\varrho_a(v, z) = \frac{v_a}{cz^\ell - v_{x(a)}}, \quad a = a(\ell, b), b \in \Pi_\ell, \quad \ell \in L \quad (55)$$

donde $v = \sum_{\omega \in W} v^\omega$. Así, $\varrho_a(v, z)$ es la relación entre el flujo promedio de abordaje de la línea ℓ en la parada b y el promedio de la capacidad disponible por servicio durante el periodo H . Implícitamente, se asume que $\varrho_a < 1$. En la relación (55), cz^ℓ es la capacidad total máxima de la línea durante el periodo H , siendo c la máxima capacidad media de los autobuses.

El tiempo promedio de cola $\varphi_a(v, z)$ por pasajero esperando en la parada vendrá dado por

$$\varphi_a(v, z) = \frac{H}{2\tau^\ell} (1 + \chi_a^2) \xi_a(\varrho_a) \quad (56)$$

donde $\tau^\ell = \max \left\{ \frac{n}{c^\ell} \sum_{\omega \in W} g_\omega, H/h_{max} \right\}, \ell \in L$.

Por consiguiente, si $a = a(\ell, b)$, el tiempo total que los pasajeros esperan para abordar una línea, ζ_a , será expresado como

$$\zeta_a = v_a \varphi_a(v, z) = v_a P_a \xi_a \left(\frac{v_a}{cz^\ell - v_{x(a)}} \right) \quad (57)$$

Si la función de demora $\zeta_a(\cdot)$ es no decreciente y convexa en $[0, 1]$, entonces la función φ_a es convexa en la región de \mathbb{R}^2 dada por $v_a + v_{x(a)} \leq cz^\ell, v_a \geq 0, v_{x(a)} \geq 0$ y por lo tanto se puede aproximar mediante un conjunto finito de planos tangentes a $(\varphi(v_a, v_{x(a)}), v_a, v_{x(a)})$ en \mathbb{R}^3 .

Si $\varphi_a^{(k)} = \varphi(v_a^{(k)}, 0)$, $\hat{\varphi}_a^{(k)} = \nabla_{v_a} \varphi(v_a^{(k)}, 0)$, $\varphi_{x(a)}^{(k)} = \nabla_{v_{x(a)}} \varphi(v_a^{(k)}, 0)$ con $(v_a^{(k)}, 0) \in S^1$ entonces la aproximación de φ en una malla de puntos convenientes puede hacerse mediante la función ψ , también convexa:

$$\psi(v_a, v_{x(a)}) = \max_{0 \leq k \leq n_\varphi - 1} \left\{ \varphi_a^{(k)} + \hat{\varphi}_a^{(k)}(v_a - v_a^{(k)}) + \hat{\varphi}_{x(a)}^{(k)} v_{x(a)} \right\} \quad (58)$$

Las demoras de los pasajeros en el modelo se incluirán de la siguiente forma al conjunto de restricciones:

$$P_a(-c\tilde{\beta}_k z^\ell + \tilde{\beta}_k v_{x(a)} + \tilde{\gamma}_k v_a) \leq \zeta_a, \quad 0 \leq k \leq n_\varphi - 1 \quad (59)$$

o equivalentemente,

$$\begin{aligned} \tilde{\beta}_{a,k} v_{x(a)} + \tilde{\gamma}_{a,k} v_a &\leq \frac{\zeta_a}{P_a} + c^\ell \tilde{\beta}_{a,k} z^\ell, \\ a &= a(\ell, b), b \in \Pi_\ell, \ell \in L, 0 \leq k \leq n_\varphi - 1 \end{aligned} \quad (60)$$

Los coeficientes $\tilde{\beta}_{a,k}, \tilde{\gamma}_{a,k}$ en la anterior restricción son:

$$\begin{aligned} \tilde{\beta}_{a,k} &= \varrho_{a,k}^2 \xi'(\varrho_{a,k}) \\ \tilde{\gamma}_{a,k} &= \xi(\varrho_{a,k}) + \varrho_{a,k} \xi'(\varrho_{a,k}) \\ a &= a(\ell, b), b \in \Pi_\ell, \ell \in L, 0 \leq k \leq n_\varphi - 1 \end{aligned} \quad (61)$$

Mediante simulaciones se ha calculado una tabla de coeficientes $\tilde{\beta}_{a,k}, \tilde{\gamma}_{a,k}$ a través de aproximaciones en diferencias finitas, la cual se encuentra en el Apéndice [A](#)

3.3.3. Modelado de capacidad finita para pasajeros esperando en una parada de autobús

También se debe tener en cuenta la capacidad limitada de las paradas para acomodar las colas de los pasajeros que esperan abordar los autobuses. Supongamos que \hat{N}_b^{pax} es la cantidad máxima de pasajeros que la parada de autobús $b \in \hat{N}_G$ puede acomodar. Se supone que puede haber una cola de pasajeros para cada una de las líneas que se detienen en la parada. La suma de las longitudes de cola correspondientes no puede exceder la capacidad \hat{N}_b^{pax} de la parada, al menos durante una gran fracción α del tiempo (digamos $\alpha = 0.95$). La aplicación de la fórmula de Little a la cola de los pasajeros que esperan autobuses de la línea ℓ en la parada de autobús $b \in \Pi_\ell$ lleva a la siguiente fórmula:

$$\frac{v_a}{H} \xi_a = \text{n}^\circ \text{ pax. medio en cola asociado al arco } a, a = a(b, \ell), b \in \Pi_\ell, \ell \in L \quad (62)$$

Sea η_b la relación entre la longitud de la cola excedida en solo una fracción $1 - \alpha$ de las veces y la longitud promedio de la cola para la parada del bus $b \in \hat{N}_G$. El modelo

$$^1S = \left\{ (v_a^{(k)}, 0) \in \mathbb{R}^2 | v_a^{(k)} \in [0, cz^\ell[, 0 \leq k \leq n_\varphi - 1 \right\}$$

requiere que la suma de todas las longitudes de cola no exceda la capacidad de una parada \hat{N}_b^{pax} durante una fracción α del tiempo bajo las siguientes restricciones:

$$\sum_{\ell \in L} \zeta_{a(\ell, b)} \leq \frac{H}{\eta_b} \hat{N}_b^{pax}, \quad b \in \hat{N}_G \quad (63)$$

3.4. Problema de programación matemática

Los tiempos en los arcos del grafo expandido vendrán dados por las funciones T_a . Los tiempos totales de espera en las colas de pasajeros en las estaciones se suponen que vendrán dados por funciones $\zeta_{a(\ell, b)}$. El tiempo total de viaje sobre la red vendrá dado por:

$$\sum_{a \in A} v_a T_a + \sum_{\ell \in L} \sum_{b \in \Pi_\ell} \zeta_{a(b, \ell)} \quad (64)$$

El primer término en (64) se corresponde con todos los tiempos en los arcos del grafo expandido con la exclusión de los tiempos de espera en cola de los pasajeros en las estaciones, componente que aparece en el segundo término.

La función objetivo del modelo consistirá en dos términos: el primero comprenderá los costes de operación de los servicios z^ℓ a la línea $\ell \in L$ más los costes de alquiler correspondientes a asignar n^ℓ buses a la línea $\ell \in L$; el segundo término se corresponderá con una valoración económica del tiempo total de los usuarios de θ €/u. de tiempo.

El modelo formulado a continuación corresponde a un problema de programación matemática lineal entero-mixto. La simplificación aplicada al modelo presentado en [8] es que las variables que deberían depender del volumen de pasajeros en los arcos del grafo v y la cantidad de servicios asignados z , se han fijado a priori, es decir, son constantes, tales como los ciclos de línea C_ℓ , el tiempo total en los arcos T_a , el tiempo por pasajero y servicio en cada parada P_a , el máximo número de servicios permitidos en la parada \hat{Z}_b , el tiempo de servicio en las paradas κ_b y el tiempo en cola en el espacio L0 de los autobuses en las paradas w_b^0 .

Las restricciones en el modelo M0 se han derivado en subsecciones anteriores y se explicarán brevemente aquí. Las restricciones A1 y A2 (véanse (41) y (47)) relacionan la disponibilidad de unidades de autobús, su asignación a líneas de autobús y el número de servicios que la flota asignada a una línea debe realizar en esa línea. Las restricciones A3 y A4 se derivan en la sección 3.3.1.4 (ver (52)), y su propósito es que, si una o más unidades de autobús están asignadas a una línea, entonces las frecuencias en esa línea deben estar dentro del intervalo $[H/h_{max}, \hat{f}]$. La factibilidad de los flujos de pasajeros según las demandas en el grafo expandido se expresa mediante la restricción B1 (ver (40)). La restricción Qb1, como se deriva en la sección 3.3.1.3 (ver (51)), es una limitación en el flujo de entrada total que una parada de autobús puede admitir para no superar la relación de servicios que permite el tiempo de servicio de autobús. La restricción Qp1 expresa la limitación impuesta por la capacidad total de la línea cz^ℓ y el flujo de pasajeros en esa línea en una estación. La restricción Qp2, como se deriva en la sección 3.3.3 (ver (63)), tiene el propósito de limitar la longitud de cola de los pasajeros que esperan en una parada

de autobús de acuerdo con el espacio disponible allí. Finalmente, la restricción Qp3, como se deriva en la sección 3.3.2 (ver (60)), corresponde a las demoras de los pasajeros.

$$\begin{aligned}
\text{[M0]} \quad & \min_{n, z, \lambda, v, \zeta} \sum_{\ell \in L} (\varsigma^\ell n^\ell + \gamma^\ell z^\ell) + \theta \sum_{a \in A} v_a T_a + \theta \sum_{\ell \in L} \sum_{b \in \Pi_\ell} \zeta_{a(\ell, b)} \\
\text{A1} \quad & \text{s.a:} \quad \sum_{\ell \in L} n^\ell \leq p \\
\text{A2} \quad & n^\ell H^\ell \geq z^\ell C_\ell, \quad \ell \in L \\
\text{A3} \quad & z^\ell \leq \lambda^\ell H \hat{f}, \quad \ell \in L \\
\text{A4} \quad & z^\ell \geq \lambda^\ell \frac{H}{h_{max}}, \quad \ell \in L \\
\text{B1} \quad & v \in \mathcal{V} \\
\text{Qb1} \quad & \sum_{\ell \in L_b} z^\ell \leq \hat{Z}_b, \quad b \in \hat{N}_G \\
\text{Qp1} \quad & v_a + v_{x(a)} \leq c z^\ell \\
\text{Qp2} \quad & \sum_{\ell \in L} \zeta_{a(\ell, b)} \leq \frac{H}{\eta_b} \hat{N}_b^{pax} \\
\text{Qp3} \quad & \tilde{\beta}_{a, k} v_{x(a)} + \tilde{\gamma}_{a, k} v_a \leq \frac{\zeta_a}{P_a} + c^\ell \tilde{\beta}_{a, k} z^\ell, \quad 0 \leq k \leq n_\varphi - 1 \\
\text{Qp4} \quad & \zeta_a \geq 0 \\
& a = a(\ell, b), b \in \Pi_\ell, \ell \in L \text{ en Qp1 a Qp4} \\
\text{Def. var.} \quad & n^\ell \in \mathbb{Z}^+, z^\ell \in \mathbb{Z}^+, \lambda^\ell \in \{0, 1\}, \ell \in L
\end{aligned}$$

Capítulo 4

Aplicación de los métodos de descomposición

En este capítulo se aplican los métodos estudiados en el capítulo 2 para resolver el modelo M0 presentado en el capítulo 3.

4.1. Modelo relajado

En el problema de programación M0 las restricciones difíciles (o complicadas) corresponden a las restricciones de capacidad Qp1 y de demora de los pasajeros Qp3, dado que dependen tanto del flujo de pasajeros, que es una variable continua, como de la cantidad de servicios, que es una variable entera. Los multiplicadores de Lagrange para Qp1 y Qp3 serán $\hat{\mu}_a^\ell$ y $\tilde{\mu}_{a,k}^\ell$, $a = a(\ell, b)$, $b \in \Pi_\ell$, $\ell \in L$, $0 \leq k \leq n_\varphi - 1$, respectivamente. Es por esto, que el problema relajado, a través del LR, queda de la siguiente forma.

$$\begin{aligned}
 [\text{M0}_{\text{LR}}] \quad \min_{\substack{n, z, \lambda \\ v, \zeta, \hat{\mu}, \tilde{\mu}}} w(\hat{\mu}_a^\ell, \tilde{\mu}_{a,k}^\ell) &= \sum_{\ell \in L} (\zeta^\ell n^\ell + \gamma^\ell z^\ell) + \theta \sum_{a \in A} v_a T_a + \theta \sum_{\ell \in L} \sum_{b \in \Pi_\ell} \zeta_{a(\ell, b)} \\
 &- \sum_{\ell \in L} \sum_{\substack{a=a(\ell, b), \\ b \in \Pi_\ell}} \hat{\mu}_a^\ell (c^\ell z^\ell - v_a - v_{x(a)}) \\
 &- \sum_{\ell \in L} \sum_{\substack{a=a(\ell, b), \\ b \in \Pi_\ell}} \sum_{0 \leq k \leq n_\varphi - 1} \tilde{\mu}_{a,k}^\ell \left(\frac{\zeta_a}{P_a} + c^\ell \tilde{\beta}_{a,k} z^\ell \right) \\
 &+ \sum_{\ell \in L} \sum_{\substack{a=a(\ell, b), \\ b \in \Pi_\ell}} \sum_{0 \leq k \leq n_\varphi - 1} \tilde{\mu}_{a,k}^\ell \left(\tilde{\beta}_{a,k} v_{x(a)} + \tilde{\gamma}_{a,k} v_a \right)
 \end{aligned}$$

$$\begin{aligned}
 \text{s.a:} \quad & \text{A1, A2, A3, A4, B1, Qb1, Qp2, Qp4 en M0} \\
 & n^\ell \in \mathbb{Z}^+, z^\ell \in \mathbb{Z}^+, \lambda^\ell \in \{0, 1\}, \ell \in L \\
 & \hat{\mu}_a^\ell \geq 0, \tilde{\mu}_{a,k}^\ell \geq 0, a = a(\ell, b), b \in \Pi_\ell, \ell \in L, 0 \leq k \leq n_\varphi - 1
 \end{aligned}$$

Se puede observar que el modelo $M0_{LR}$ es separable, es decir, se puede dividir en subproblemas. En este caso, se puede dividir en un subproblema entero LR_z que contenga a las variables enteras n^ℓ , z^ℓ y λ^ℓ , y en dos subproblemas continuos LR_v y LR_ζ , que contengan las variables de flujo v_a y las variables de tiempo de espera ζ_a , respectivamente, como se muestra a continuación.

$$\begin{aligned}
[LR_z] \quad & \min_{n,z,\lambda} \quad \sum_{\ell \in L} (\varsigma^\ell n^\ell + \gamma^\ell z^\ell) \\
& - \sum_{\ell \in L} \sum_{\substack{a=a(\ell,b), \\ b \in \Pi_\ell}} \hat{\mu}_a^{\ell(j)} c^\ell z^\ell \\
& - \sum_{\ell \in L} \sum_{\substack{a=a(\ell,b), \\ b \in \Pi_\ell}} \sum_{0 \leq k \leq n_\varphi - 1} \tilde{\mu}_{a,k}^{\ell(j)} c^\ell \tilde{\beta}_{a,k} z^\ell \\
\text{s.a:} \quad & A1, A2, A3, A4, Qb1 \text{ en } M0
\end{aligned}$$

$$\begin{aligned}
[LR_v] \quad & \min_v \quad \theta \sum_{a \in A} v_a T_a \\
& + \sum_{\ell \in L} \sum_{\substack{a=a(\ell,b), \\ b \in \Pi_\ell}} \hat{\mu}_a^{\ell(j)} (v_a + v_{x(a)}) \\
& + \sum_{\ell \in L} \sum_{\substack{a=a(\ell,b), \\ b \in \Pi_\ell}} \sum_{0 \leq k \leq n_\varphi - 1} \tilde{\mu}_{a,k}^{\ell(j)} \left(\tilde{\beta}_{a,k} v_{x(a)} + \tilde{\gamma}_{a,k} v_a \right) \\
\text{s.a:} \quad & B1 \text{ en } M0
\end{aligned}$$

$$\begin{aligned}
[LR_\zeta] \quad & \min_{\zeta} \quad \theta \sum_{\ell \in L} \sum_{b \in \Pi_\ell} \zeta_{a(\ell,b)} \\
& - \sum_{\ell \in L} \sum_{\substack{a=a(\ell,b), \\ b \in \Pi_\ell}} \sum_{0 \leq k \leq n_\varphi - 1} \tilde{\mu}_{a,k}^{\ell(j)} \frac{\zeta_a}{P_a} \\
\text{s.a:} \quad & Qp2, Qp4 \text{ en } M0
\end{aligned}$$

$\hat{\mu}_a^{\ell(i)}$ y $\tilde{\mu}_{a,k}^{\ell(i)}$ son las i -ésimas soluciones obtenidas del proceso iterativo correspondiente y son un parámetro de entrada.

4.2. Resolución de $M0_{LR}$ mediante el método del Subgradiente

Se puede observar que, dado que los multiplicadores de Lagrange se fijan a priori, LR_v puede ser resuelto mediante el algoritmo de *Camino mínimo*.

Habitualmente se les da valor cero a los multiplicadores de Lagrange, aunque a veces puede ser conveniente asignar un valor relacionado con el problema a resolver. En este caso, se inicializarán en cero.

El algoritmo para resolver el modelo mediante el método del subgradiente se presenta a continuación.

Algoritmo 3 Resolver $M0_{LR}$ mediante el Método del Subgradiente

- 1: Hacer $j = 0$, $\hat{\mu}_a^{\ell,(j)} = 0$ y $\tilde{\mu}_{a,k}^{\ell,(j)} = 0$, $a = a(\ell, b)$, $b \in \Pi_\ell$, $\ell \in L$, $0 \leq k \leq n_\varphi - 1$
- 2: Resolver MP_{M0} y obtener $Z_{mp}^{(j)}$, $\hat{\mu}_a^{\ell,(j)}$ y $\tilde{\mu}_{a,k}^{\ell,(j)}$, $a = a(\ell, b)$, $b \in \Pi_\ell$, $\ell \in L$, $0 \leq k \leq n_\varphi - 1$
- 3: Resolver LR_z y obtener $n^{\ell,(j+1)}$, $z^{\ell,(j+1)}$, $\lambda^{\ell,(j+1)}$, $\ell \in L$
- 4: Resolver LR_ζ y obtener $\zeta_{a(\ell,b)}^{(j+1)}$, $b \in \Pi_\ell$, $\ell \in L$
- 5: Resolver LR_v mediante “Camino mínimo” y obtener $v_a^{(j+1)}$, $a \in A$
- 6: Calcular los subgradientes de $Qp1$ y $Qp3$, \hat{g} y \tilde{g} respectivamente:

$$\hat{g}_a^{\ell,(j)} = z^\ell c^\ell - v_a - v_{x(a)} \quad (65)$$

$$\tilde{g}_{a,k}^{\ell,(j)} = \frac{\zeta_a}{P_a} + c^\ell \tilde{\beta}_{a,k} z^\ell - \tilde{\beta}_{a,k} v_{x(a)} - \tilde{\gamma}_{a,k} v_a \quad (66)$$

$$a = a(\ell, b), b \in \Pi_\ell, \ell \in L, 0 \leq k \leq n_\varphi - 1$$

- 7: Si $relgap^{(j)} \leq \varepsilon$ entonces STOP, donde

$$gap^{(j)} = \sum_{\ell \in L} \sum_{\substack{a=a(\ell,b), \\ b \in \Pi_\ell}} \hat{\mu}_a^{\ell,(j)} \hat{g}_a^{\ell,(j)} + \sum_{\ell \in L} \sum_{\substack{a=a(\ell,b), \\ b \in \Pi_\ell}} \sum_{0 \leq k \leq n_\varphi - 1} \tilde{\mu}_{a,k}^{\ell,(j)} \tilde{g}_{a,k}^{\ell,(j)} \quad (67)$$

$$relgap^{(j)} = \frac{gap^{(j)}}{\|(\hat{\mu}_a^{\ell,(j)}, \tilde{\mu}_{a,k}^{\ell,(j)})\| \|(\hat{g}_a^{\ell,(j)}, \tilde{g}_{a,k}^{\ell,(j)})\| + \epsilon} \quad (68)$$

- 8: Calcular la longitud de paso $\alpha^{(j)}$
 - 9: Calcular $\hat{\mu}_a^{\ell,(j+1)} = \max \{0, \hat{\mu}_a^{\ell,(j)} - \alpha^{(j)} \hat{g}_a^{\ell,(j)}\}$
 - 10: Calcular $\tilde{\mu}_{a,k}^{\ell,(j+1)} = \max \{0, \tilde{\mu}_{a,k}^{\ell,(j)} - \alpha^{(j)} \tilde{g}_{a,k}^{\ell,(j)}\}$
 - 11: Hacer $j = j + 1$ y volver al paso 2.
-

Ambos métodos planteados en [14](#) y mencionados en la sección [2.3.1](#) serán utilizados para calcular la longitud de paso en el paso 7.

4.3. Resolución de $M0_{LR}$ mediante el algoritmo de Cutting Plane

El SP de $M0_{LR}$ corresponde a los subproblemas LR_z , LR_v y LR_ζ . Al igual que en el método de subgradiente, el problema LR_v se resolverá por el algoritmo de camino

mínimo. El MP queda expresado de la siguiente forma, donde $n^{(i)}, z^{(i)}, \lambda^{(i)}, v^{(i)}$ y $\zeta^{(i)}$ corresponden a la solución del i -ésimo SP y j es el número de cortes.

$$\begin{aligned}
[\text{MP}_{\text{M0}}] \quad & \max_{x, \mu, \tilde{\mu}} \quad Z_{mp} \\
s.a \quad & Z_{mp} \leq \sum_{\ell \in L} (\zeta^\ell n^{\ell(i)} + \gamma^\ell z^{\ell(i)} + \theta \sum_{a \in A} v_a^{(i)} T_a + \theta \sum_{\ell \in L} \sum_{b \in \Pi_\ell} \zeta_{a(\ell, b)}^{(i)} \\
& - \sum_{\ell \in L} \sum_{\substack{a=a(\ell, b), \\ b \in \Pi_\ell}} \hat{\mu}_a^\ell (c^\ell z^{\ell(i)} - v_a^{(i)} - v_{x(a)}^{(i)}) \\
& - \sum_{\ell \in L} \sum_{\substack{a=a(\ell, b), \\ b \in \Pi_\ell}} \sum_{0 \leq k \leq n_\varphi - 1} \tilde{\mu}_{a, k}^\ell \left(\frac{\zeta_a^{(i)}}{P_a} + c^\ell \tilde{\beta}_{a, k} z^{\ell(i)} - \tilde{\beta}_{a, k} v_{x(a)}^{(i)} - \tilde{\gamma}_{a, k} v_a^{(i)} \right) \\
& i = 0, \dots, j \\
& \hat{\mu}_a^\ell \geq 0, \tilde{\mu}_{a, k}^\ell \geq 0, a = a(\ell, b), b \in \Pi_\ell, \ell \in L, 0 \leq k \leq n_\varphi - 1
\end{aligned}$$

El algoritmo de Cutting Plane de Dantzig para resolver M0_{LR} queda de la siguiente forma.

Algoritmo 4 Resolver M0_{LR} mediante el algoritmo Cutting Plane de Dantzig

- 1: Hacer $j = 0$. Encontrar una solución factible inicial $n^{(0)}, z^{(0)}, \lambda^{(0)}, v^{(0)}, \zeta^{(0)}$
 - 2: Resolver MP_{M0} y obtener $Z_{mp}^{(j)}, \hat{\mu}_a^{\ell(j)}$ y $\tilde{\mu}_{a, k}^{\ell(j)}$, $a = a(\ell, b)$, $b \in \Pi_\ell$, $\ell \in L$, $0 \leq k \leq n_\varphi - 1$
 - 3: Resolver LR_z y obtener $n^{\ell(j+1)}, z^{\ell(j+1)}, \lambda^{\ell(j+1)}$, $\ell \in L$
 - 4: Resolver LR_ζ y obtener $\zeta_{a(\ell, b)}^{(j+1)}$, $b \in \Pi_\ell$, $\ell \in L$
 - 5: Resolver LR_v mediante “Camino mínimo” y obtener $v_a^{(j+1)}$, $a \in A$
 - 6: Obtener $w(\hat{\mu}_a^{\ell(j)}, \tilde{\mu}_{a, k}^{\ell(j)})$
 - 7: **if** $\left(Z_{mp}^{(j)} \leq w(\hat{\mu}_a^{\ell(j)}, \tilde{\mu}_{a, k}^{\ell(j)}) \right) \vee \left(\frac{Z_{mp}^{(j)} - w(\hat{\mu}_a^{\ell(j)}, \tilde{\mu}_{a, k}^{\ell(j)})}{w(\hat{\mu}_a^{\ell(j)}, \tilde{\mu}_{a, k}^{\ell(j)})} \leq \varepsilon_{mp} \right)$ **then**
 - 8: STOP
 - 9: **else**
 - 10: Agregar un nuevo corte a MP_{M0}
 - 11: Hacer $j = j + 1$ e ir al paso 2
 - 12: **end if**
-

Los procedimientos utilizados para acelerar el algoritmo anterior se explicaron en la sección [2.4.3](#).

4.3.1. Obtención de una solución factible inicial

La forma más sencilla de obtener una solución factible inicial es calcular el volumen de pasajeros suponiendo que todos los desplazamientos se realizan a pie, a través de los correspondientes arcos de la red, por lo que no se necesitan autobuses y no habrá esperas en las paradas, entonces el resto de variables del modelo M0 serán cero.

4.3.2. Recuperación de la solución primal

Para recuperar la solución primal de $\boxed{\text{M0}}$ sería conveniente seguir el procedimiento explicado en la sección $\boxed{2.4.2}$. Sin embargo, este procedimiento funciona correctamente solo con problemas lineales, y dado que el modelo $\boxed{\text{M0}}$ contiene variables enteras y binarias, se ha hecho una modificación.

Lo que se hace principalmente es agregar las restricciones $\sum_{i=0}^j \alpha^{(i)} = 1$, $\alpha^{(i)} \geq 0$, donde j es el número de cortes, al modelo $\boxed{\text{M0}}$ y las variables continuas, flujos de pasajeros v y tiempos de espera ζ , son reformulados de la siguiente forma

$$\begin{aligned}\tilde{v}_a &= \sum_{i=0}^j \alpha^{(i)} \hat{v}_a^{(i)}, \quad a \in A \\ \tilde{\zeta}_{a(\ell,b)} &= \sum_{i=0}^j \alpha^{(i)} \hat{\zeta}_{a(\ell,b)}^{(i)}, \quad b \in \Pi_\ell, \ell \in L\end{aligned}$$

donde $\hat{v}^{(i)}$ y $\hat{\zeta}^{(i)}$ son las soluciones del i -ésimo SP, y por lo tanto se eliminan las restricciones B1 y Qp2. De esta forma, el nuevo modelo tendrá como variables a α, z, n y λ . El modelo a resolver queda representado en $\boxed{\text{M0}_{\text{rec}}}$

$$\begin{aligned}[\text{M0}_{\text{rec}}] \quad & \min_{\alpha, n, z, \lambda} \quad \sum_{\ell \in L} (\varsigma^\ell n^\ell + \gamma^\ell z^\ell) + \theta \sum_{a \in A} \tilde{v}_a T_a + \theta \sum_{\ell \in L} \sum_{b \in \Pi_\ell} \tilde{\zeta}_{a(\ell,b)} \\ \text{s.a:} \quad & \text{A1, A2, A3, A4, Qb1 en } \boxed{\text{M0}} \\ \text{Qp1'} \quad & \tilde{v}_a + \tilde{v}_{x(a)} \leq c z^\ell \\ \text{Qp3'} \quad & \tilde{\beta}_{a,k} \tilde{v}_{x(a)} + \tilde{\gamma}_{a,k} \tilde{v}_a \leq \frac{\tilde{\zeta}_a}{P_a} + c^\ell \tilde{\beta}_{a,k} z^\ell, \quad 0 \leq k \leq n_\varphi - 1 \\ \text{Qp4'} \quad & \tilde{\zeta}_a \geq 0 \\ \text{R1} \quad & \tilde{v}_a = \sum_{i=0}^j \alpha^{(i)} \hat{v}_a^{(i)}, \quad a \in A \\ \text{R2} \quad & \tilde{\zeta}_a = \sum_{i=0}^j \alpha^{(i)} \hat{\zeta}_a^{(i)} \\ \text{R3} \quad & \sum_{i=0}^j \alpha^{(i)} = 1, \quad \alpha^{(i)} \geq 0 \\ & a = a(\ell, b), b \in \Pi_\ell, \ell \in L \text{ en Qp1'}, \text{ Qp3'}, \text{ Qp4'} \text{ y R2} \\ \text{Def. var.} \quad & n^\ell \in \mathbb{Z}^+, z^\ell \in \mathbb{Z}^+, \lambda^\ell \in \{0, 1\}, \ell \in L\end{aligned}$$

Capítulo 5

Resultados numéricos

En este capítulo se presentan los resultados obtenidos al resolver el modelo [M0](#) por los métodos de descomposición propuestos, así como la comparación entre ellos y su efectividad.

5.1. Casos de estudio

Para ilustrar el comportamiento del modelo y los métodos propuestos para resolverlo, se han utilizado dos casos de prueba, uno de tamaño pequeño y otro de tamaño mediano, que corresponden a las mismas redes estudiadas en [8](#). Se tratan de una propuesta de diseño pensadas como sustitución en caso de una interrupción del sistema de transporte.

La red de tamaño pequeño se ubica en la ciudad de Madrid, durante la interrupción del corredor ferroviario del Paseo de La Castellana, mientras que la de tamaño mediano se ubica en Barcelona, durante la interrupción de la línea 1 del metro (desde las estaciones Plaça d'Espanya hasta Clot). En la Tabla [1](#) se presentan las características de cada una de ellas, así como el número de variables y restricciones para el modelo [M0](#).

La tabla de viajes origen-destino, con más de 37.000 pasajeros en un período punta de tres horas, para el corredor de Madrid y la línea de metro de Barcelona, aparecen en la Tablas [9](#) y [10](#) (Apéndice [B](#)), respectivamente.

La Figura [5](#) muestra una representación esquemática del diseño del Paseo de la Castellana en Madrid, donde se pretende que funcione el sistema, y la Figura [6](#) muestra un esquema de las arterias principales donde el sistema puede operar en Barcelona (Gran Vía de les Corts Catalanes, Av. Meridiana, así como algunas calles menores).

Ambas redes de autobuses expandidas contienen un conjunto de enlaces A_G (como se define en [36](#)) para el movimiento de pasajeros realizado fuera de la red de autobuses (acceso desde estaciones interrumpidas a paradas de autobús, transferencias entre líneas de autobús, partes del viaje realizadas a pie). Este conjunto de

Tabla 1. Parámetros de las redes estudiadas

Red	Líneas	Paradas	Nodos	Arcos	Pares O-D	Centroides	Viajes	Var.	Rest.
Madrid	20	6	118	240	12	4	37000	1318	1449
Barcelona	48	17	310	640	88	10	37992	7328	5599

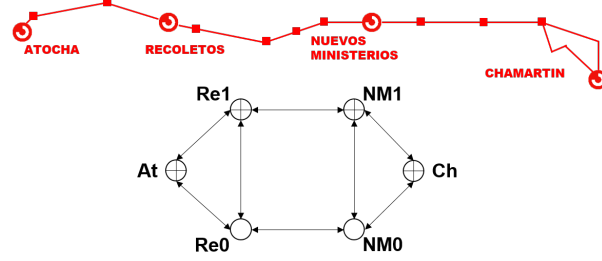


Figura 5. (Arriba) Representación esquemática de la distribución del Paseo de la Castellana en Madrid. (Abajo) Representación del subgrafo (N_G, A_G) para movimientos de pasajeros fuera de las líneas de extensión de autobuses. Las paradas de autobuses en Recoletos y Nuevos Ministerios se han reflejado a ambos lados del Paseo de la Castellana

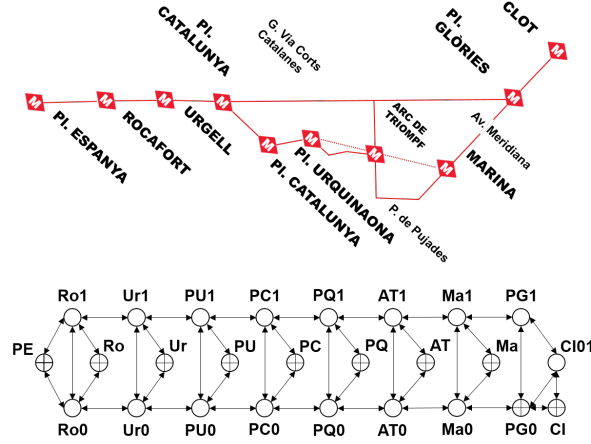


Figura 6. (Arriba) Representación esquemática de la línea 1 del metro en Barcelona. (Abajo) Representación del subgrafo (N_G, A_G) para los movimientos de pasajeros fuera de la extensión de líneas de autobús. Excluyendo Plaça d'Espanya y Clot, todas las estaciones de autobuses se han duplicado en ambos lados de la línea.

enlaces A_G se representa en la parte inferior de las Figs. 5 y 6. En ambos casos las estaciones interrumpidas se muestran como nodos marcados con una cruz.

En ambas redes de prueba, (a) se han asumido unidades de autobuses con una capacidad de 100 pasajeros; (b) el período de tiempo considerado es un período máximo de $H = 180$ minutos; (c) el tiempo de servicios en las paradas κ_b se ha establecido en 45 segundos; (d) el valor del tiempo θ se ha supuesto que es 0.081 €/min; y (e) las distancias entre paradas de autobús se han estimado de manera realista y se ha supuesto que existen carriles de autobuses apropiados para mantener los autobuses funcionando a una velocidad promedio de 26 km/h entre paradas de autobús.

5.1.1. Implementación

Los algoritmos de descomposición han sido implementados en AMPL, usando CPLEX 12.5.0.0 como solver. El algoritmo de camino mínimo se ha desarrollado en lenguaje C, mediante la plataforma Visual Studio 2010.

Las pruebas se han ejecutado en un ordenador con procesador x64 Intel(R) Xeon(R) de 2.4GHz y con 48 Gb de RAM, SO Windows Server 2012 Standard.

5.2. Resultados de aplicar el Método del Subgradiente

En la Figura 7 se muestra la convergencia del gap relativo para 3.000 iteraciones, de acuerdo a la ecuación (29), cuando se resuelve el problema M0 mediante el método del Subgradiente. Los gráficos superiores corresponden a los resultados obtenidos al actualizar los multiplicadores de Lagrange con el método del MSWA (ver (26)) y los gráficos inferiores corresponden a la actualización mediante el método de media autorregulada (ver (27) y (28)). Para el método MSWA $d = 2$ y para el método de media autorregulada el valor de los parámetros es $\Gamma = 1.9$ y $\gamma = 0.01$.

Teóricamente el gap relativo debería converger a cero, sin embargo, se puede observar en la Figura 7 que para ninguno de los dos casos ocurre. De acuerdo a [18], cuando existen óptimos alternativos el método del Subgradiente es inestable, en el sentido de que para diferentes puntos iniciales puede converger a diferentes soluciones e incluso puede no estar garantizada la convergencia. Este tipo de problemas pueden ser resueltos mediante un *método de regularización*, que sería interesante analizar en futuras investigaciones.

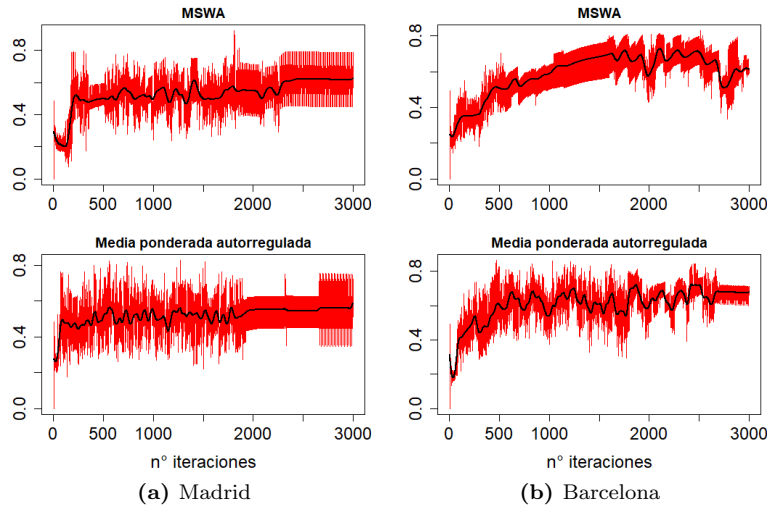


Figura 7. *relgap* obtenido mediante el método del Subgradiente para ambas redes de acuerdo al método utilizado para actualizar los multiplicadores de Lagrange

En conclusión, este método no es el adecuado para resolver el modelo **M0**, independientemente del método que se utilice para actualizar los multiplicadores de Lagrange.

5.3. Resultados de aplicar el Algoritmo de Cutting Plane de Dantzig

A continuación se presentan los resultados obtenidos al aplicar el algoritmo de Cutting Plane de Dantzig y los métodos de aceleración, explicados en la Sección 4.3, para resolver el modelo **M0** en ambas redes.

En las Tablas 2 y 3 las etiquetas de las filas corresponden a los métodos aplicados. La fila “CP” corresponde a los resultados con el algoritmo sin aceleración. La fila “CP + MSWA” corresponde a los resultados obtenidos de aplicar el método MSWA para actualizar los multiplicadores de Lagrange y acelerar el algoritmo (ver Sección 2.4.3). Por último, la fila “CP + acel.” corresponde a los resultados obtenidos al aplicar la aceleración presentada en la Figura 1 (Sección 2.4.3).

Además, en ambas tablas la columna “# iter” corresponde a la cantidad de iteraciones requeridas por el método para lograr un $relgap \leq \varepsilon_{mp}$. La columna “T. ejec.” corresponde al tiempo de ejecución de todo el proceso iterativo. Las columnas “ Z_{mp} ” y “ $w(\mu)$ ” muestran el valor de la función objetivo del MP y del SP en la última iteración, respectivamente. La columna “relgap” muestra el gap relativo alcanzado en la última iteración, de acuerdo a la Eq. (31). Las columnas “ $f^{(k)}$ ” y “ f_{rec}^* ” presentan el valor de la función objetivo de **M0** en la última iteración y en la recuperación de la solución primal, respectivamente. Por último, la columna “% a pie” muestra el porcentaje de viajes que se realizan a pie de acuerdo a los resultados de la recuperación.

La tolerancia exigida ε_{mp} para el gap relativo es 0.01 y 0.05, para el caso de Madrid y Barcelona, respectivamente. Con respecto a la solución factible inicial (ver Sección 4.3.1), el costo total por enviar a todos los pasajeros a pie es $f^0 = 237323.68$ y $f^0 = 150330.61$, respectivamente. Para ambos casos en el método de aceleración $K = 40$ y $S = 3$ y en el método MSWA $d = 2$. Además, se estableció un límite de 2000 iteraciones en todas las ejecuciones.

En los resultados presentados en las Tablas 2 y 3 se puede observar que en ambos casos la aplicación del método MSWA para actualizar los multiplicadores de Lagrange es el más eficiente, dado que es el único que realmente acelera el algoritmo de Cutting Plane. Además, la solución final f_{rec}^* no se ve demasiado afectada, con respecto al resultado del algoritmo sin aceleración, y su tiempo de ejecución es razonable.

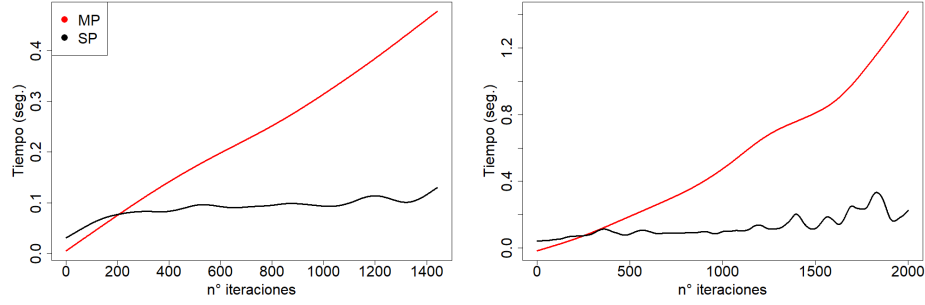
Con respecto a los resultados con la aplicación del método de aceleración, para la red de Madrid se puede apreciar que se logra la condición del $relgap < 0.01$ a casi 500 iteraciones más que sin utilizar ningún método de aceleración, obteniendo resultados similares. Además, para la red de Barcelona no se alcanza un $relgap < 0.05$ antes de las 2000 iteraciones, aunque el resultado en ese punto es similar a los obtenidos sin

Tabla 2. Resultados de resolver el modelo **M0** mediante Cutting Plane y los métodos de aceleración para la red de Madrid

Algoritmo	# iter.	T. ejec.	Z_{mp}	$w(\mu)$	$relgap$	f^k	f_{rec}^*	% a pie
CP	937	24 min.	167518.60	166067.30	0.0087	159820.21	175897.29	56.84
CP + MSA	644	11 min.	167538.39	165983.94	0.0094	147844.44	176321.89	56.61
CP + acel.	1441	1.33 h.	167608.09	165956.78	0.01	138523.12	176150.79	57.92

Tabla 3. Resultados de resolver el modelo **M0** mediante Cutting Plane y los métodos de aceleración para la red de Barcelona

Algoritmo	# iter.	T. ejec.	Z_{mp}	$w(\mu)$	$relgap$	f^k	f_{rec}^*	% a pie
CP	1598	4.5 h	102412.97	97806.21	0.0471	122372.28	103866.20	39.47
CP + MSA	878	1.33 h	102591.52	97754.18	0.0495	138382.84	104573.30	39.32
CP + acel.	2000	11.33 h	102513.02	91894.26	0.1156	159148.92	103876.12	39.78

**Figura 8.** Tiempo de ejecución (seg.) del MP y SP para el método de aceleración del Cutting Plane para las redes de Madrid (izquierda) y Barcelona (derecha)

método de aceleración y con MSA. Analizando los resultados, el mínimo $relgap$ alcanzado para la red de Barcelona fue a las 1875 iteraciones, con un $relgap = 0.051$ y una solución final $f_{rec}^* = 103968.07$, por lo que no hay una importante mejora con más iteraciones. Pese a que el método intenta reducir la cantidad de cortes en ciertas iteraciones y el tiempo de resolución del MP, el MP puede llegar a resolverse hasta cuatro veces, por lo que el tiempo de ejecución vuelve a aumentar. En la Figura 8 se muestra este fenómeno.

La lentitud de la resolución mediante el método de aceleración también se puede evidenciar en las Figuras 9 y 10, dado que los valores del MP y del SP son muy inestables a lo largo de todo el proceso iterativo, así como el $relgap$.

En la Figura 9 se muestra la convergencia de los tres métodos para ambos casos de estudio. Se puede ver que para la red de Madrid sobre las 300 iteraciones los valores Z_{mp} y $w(\mu)$ comienzan a ser muy cercanos, pero la tendencia a ser iguales es lenta. En cambio, para la red de Barcelona el gap entre estos valores se reduce muy lentamente en cada iteración, e incluso para los casos de resolución sin aceleración y con MSA, el valor del SP tiene una importante caída en las últimas iteraciones, pero se recupera rápidamente.

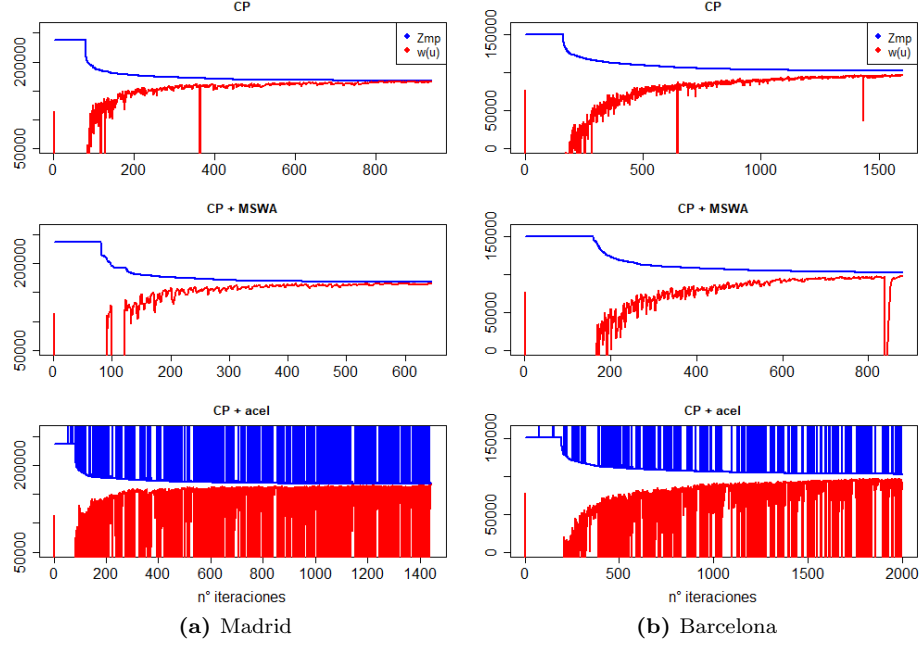


Figura 9. Convergencia del MP (Z_{mp}) y del SP ($w(\mu)$) para ambas redes mediante los métodos de resolución CP: Cutting Plane puro; CP+MSWA: Cutting Plane más actualización de multiplicadores por el método MSWA; CP+acel: Cutting Plane más aceleración

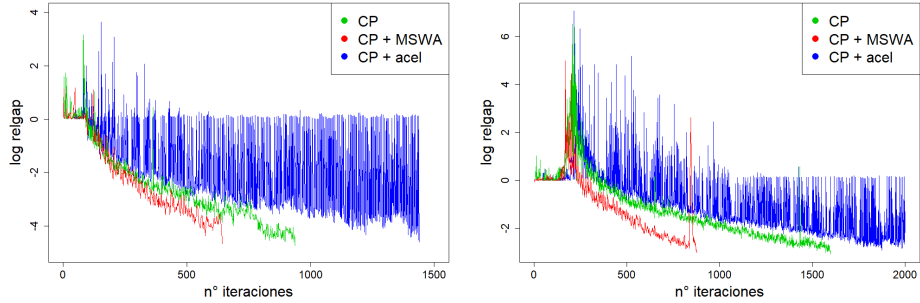


Figura 10. $relgap$ (en escala logarítmica) para la red de Madrid (izquierda) y Barcelona (derecha) mediante los métodos de resolución CP: Cutting Plane puro; CP+MSWA: Cutting Plane más actualización de multiplicadores por el método MSWA; CP+acel: Cutting Plane más aceleración

La Figura 10 complementa los análisis anteriores de que, tanto para la red de Madrid como para la de Barcelona, el mejor método de resolución para el modelo M0 es aplicando la actualización de los multiplicadores de Lagrange con el método MSWA al Cutting Plane. Por esto que en las Tablas 4 y 5 se presentan los resultados obtenidos con este método.

En la Tabla 4 la primera columna “Línea ℓ ” corresponde a las líneas seleccionadas para operar en el sistema de extensión. Las columnas “From” y “To” indican el origen y destino de la línea ℓ , respectivamente. Las columnas “T. Cola” denotan el tiempo medio de espera de los pasajeros hasta que toman un autobús de la línea

ℓ en las estación **From** y **To**, respectivamente. Las columnas “**Occ-**” y “**Occ+**” denotan la relación entre el flujo de pasajeros y la capacidad de pasajeros en línea en dirección **From** a **To** y en la dirección opuesta, respectivamente. “ n^ℓ ” indica el número de vehículos asignados y “ z^ℓ ” el número de servicios llevados a cabo por la línea ℓ durante el horizonte de tiempo. “ C_ℓ ” indica la duración del ciclo y por último “**head.**” indica el tiempo promedio entre llegadas de autobuses en una parada de la línea. Por brevedad, solo se presentan las últimas cuatro columnas para la red de Barcelona en el la tabla 5, dado que de 48 líneas, 43 son seleccionadas.

Para la red de Madrid, de 20 líneas posibles, el método con MSWA selecciona 11, aunque solo tres de ellas (L4A, L8 y L9) tienen un importante número de vehículos y servicios asignados y con los menores headway. Se observa además, que varias de las líneas seleccionadas tienen el mismo (o similar) recorrido, por lo que esta solución no es práctica en términos de implementación. Esto ocurre porque la recuperación de la solución primal considera todas las soluciones del proceso iterativo.

En el caso de la red de Barcelona, la solución tampoco es práctica, dado que se selecciona casi el 90 % de las líneas candidatas, de las cuales a 20 líneas se le asigna un vehículo y a 10 dos vehículos, y en la mayoría de los casos la cantidad de servicios no supera los cinco. Y ocurre lo mismo que para Madrid, muchos recorridos se repiten. Por lo que quizás sería necesario aplicar una nueva restricción que limite la mínima cantidad de vehículos a contratar por el sistema, para cada línea.

Analizando la Tabla 4 para la red de Madrid, se observa que las capacidades de línea son suficientes para acomodar a los pasajeros e incluso en algunos casos el porcentaje de ocupación es bajo. Sin embargo, la espera por un autobús apropiado es relevante, que al parecer se debe al elevado tiempo entre llegada de autobuses. En el caso de Barcelona, el porcentaje de ocupación no suele superar el 50 % y los tiempos en la cola de espera son en promedio 5 minutos, debido a que existen muchas líneas que hacen el mismo recorrido.

Por último, en las Tablas 6 y 7 se muestra la cantidad de servicios total que operan, la cantidad de servicios máxima admitida \hat{Z}_b , el porcentaje de ocupación de servicios y el factor de carga (ver (44)) de cada parada $b \in \hat{N}_G$, para Madrid y Barcelona respectivamente. Se puede apreciar que para el caso de Madrid, ninguna estación atiende el máximo de servicios posibles y los factores de carga ρ_b son relativamente bajos. En cambio para el caso de Barcelona, el porcentaje de servicios que se atienden en las paradas, respecto del máximo posible, es alto, en promedio un 84 %, que puede deberse al hecho de que la solución de recuperación selecciona muchas líneas con el mismo (o similar) recorrido y por lo tanto, las paradas se ven saturadas de servicios.

Tabla 4. Líneas de autobuses seleccionadas por el método de Cutting Plane más el método MSWA y sus parámetros para la red de Madrid

Línea ℓ	From	T. Cola (min.)	Occ+	To	T. Cola (min.)	Occ-	n^ℓ	z^ℓ	C_ℓ	head. (min.)
L1B	At	7.95	69.02	Re1	4.44	15.37	2	9	37	20
L3B	NM1	4.43	6.02	Ch	4.44	15.12	3	11	49	16.36
L4A	At	11.69	81.23	Re1	4.43	5.14	18	38	85	4.74
L4B	At	8.22	70.82	NM1	12.72	82.24	5	10	85	18
L5B	At	8.22	70.82	NM1	7.15	62.21	5	10	83	18
L6A	Re0	7.01	63.31	Ch	5.36	17.63	5	10	88	18
L6B	Re1	8.22	70.82	Ch	5.53	50.51	5	10	88	18
L8	At	10.27	76.61	Ch	6.44	59.86	18	28	115	6.43
L9	At	11.69	81.23	Ch	5.53	50.51	27	43	113	4.19
L10	At	10.89	78.56	Ch	4.57	28.18	6	9	112	20
L11	At	4.7	32.86	Ch	33.63	15.86	6	9	109	20

Tabla 5. Líneas de autobuses seleccionadas por el método de Cutting Plane más el método MSWA y sus parámetros para la red de Barcelona

Línea ℓ	From	To	n^ℓ	z^ℓ	C_ℓ	head. (min.)	Línea ℓ	From	To	n^ℓ	z^ℓ	C_ℓ	head. (min.)
L1S1	PE	Ur1	4	17	41.5	10.59	L11ES	PE	Cl01	2	4	77.3	45
L2S1	PU0	PC0	1	6	29.6	30	L1EAAP	PE	AT1	4	13	54.7	13.85
L3S1	PE	PC0	4	13	52.8	13.85	L1EATP	PE	AT1	3	10	53.3	18
L1S2	PC0	PQ0	1	7	23.5	25.71	L2EAAP	PC0	Cl01	2	4	65.5	45
L3S2	PC0	AT1	1	6	27.4	30	L2EATP	PC0	Cl01	2	5	63.9	36
L1S3	AT1	Ma1	1	6	29.1	30	L3EAAP	PE	Cl01	4	8	85.4	22.5
L4S3	AT1	PG1	1	3	48.5	60	L3EATP	PE	Cl01	3	6	86.7	30
L5S3	AT1	PG1	1	3	47.5	60	L4EAAP	PE	Cl01	3	6	84.3	30
L6S3	Ma1	Cl01	1	4	42.2	45	L4EATP	PE	Cl01	2	4	87	45
L7S3	Ma1	Cl01	1	3	36.5	60	L5EAAP	PE	Cl01	3	6	82.3	30
L8S3	AT1	Cl01	5	15	60	12	L5EATP	PE	Cl01	2	4	87	45
L9S3	AT1	Cl01	2	6	58	30	L1EAA	PE	AT1	1	4	26.95	45
L10S3	AT1	Cl01	2	6	54.3	30	L1EAT	PE	AT1	2	13	25.95	13.85
L11S3	AT1	Cl01	1	3	52.3	60	L2EAA	PC0	Cl01	1	6	26.65	30
L1ES	PE	PC0	1	3	48.2	60	L2EAT	PC0	Cl01	1	3	38.85	60
L2ES	PC0	AT1	1	6	26.6	30	L3EAA	PE	Cl01	1	4	44.35	45
L4ES	PE	AT1	7	23	54.4	7.83	L3EAT	PE	Cl01	1	3	48.95	60
L5ES	PE	AT1	2	6	53.6	30	L4EAA	PE	Cl01	1	3	43.25	60
L6ES	PC0	Cl01	2	5	67	36	L4EAT	PE	Cl01	1	3	49.25	60
L8ES	PE	Cl01	12	22	94.8	8.18	L5EAA	PE	Cl01	1	4	41.25	45
L9ES	PE	Cl01	4	7	94	25.71	L5EAT	PE	Cl01	1	3	49.25	60
L10ES	PE	Cl01	4	7	92	25.71							

Tabla 6. Cantidad de servicios que atiende cada parada con respecto al máximo que puede admitir, para la red de Madrid

Parada b	$\sum_{\ell \in L_b} z^\ell$	\hat{Z}_b	$\sum_{\ell \in L_b} z^\ell / \hat{Z}_b$	ρ_b
At	156	240	65 %	0.0406
Ch	120	240	50 %	0.1000
NM0	178	480	37.08 %	0.0927
NM1	178	240	74.17 %	0.1060
Re0	176	320	55 %	0.1833
Re1	176	240	73.33 %	0.1467

Tabla 7. Cantidad de servicios que atiende cada parada con respecto al máximo que puede admitir, para la red de Barcelona

Parada b	$\sum_{\ell \in L_b} z^\ell$	\hat{Z}_b	$\sum_{\ell \in L_b} z^\ell / \hat{Z}_b$	ρ_b	Parada b	$\sum_{\ell \in L_b} z^\ell$	\hat{Z}_b	$\sum_{\ell \in L_b} z^\ell / \hat{Z}_b$	ρ_b
AT0	234	240	97.50 %	0.4875	PG1	160	240	66.67 %	0.1667
AT1	240	240	100.00 %	0.2	PQ0	205	240	85.42 %	0.0854
CI01	154	240	64.17 %	0.1069	PQ1	205	240	85.42 %	0.1708
Ma0	160	240	66.67 %	0.1667	PU0	202	240	84.17 %	0.2104
Ma1	166	240	69.17 %	0.0865	PU1	202	240	84.17 %	0.2104
PC0	227	240	94.58 %	0.2365	Ro0	196	240	81.67 %	0.1361
PC1	227	240	94.58 %	0.1051	Ro1	196	240	81.67 %	0.1361
PE	196	240	81.67 %	0.1633	Ur0	196	240	81.67 %	0.1361
PG0	160	160	100.00 %	0.6667	Ur1	196	240	81.67 %	0.1361

Capítulo 6

Conclusiones y futuras investigaciones

En este trabajo se han propuesto diferentes métodos de resolución para la selección de líneas candidatas en extensiones de transporte público, en base al método del Subgradiente y el algoritmo de Cutting Plane de Dantzig. Este problema fue desarrollado inicialmente en [8] y aquí se ha realizado una simplificación, dado que se consideran casos de congestión mediana.

Para ambos casos estudiados, Madrid y Barcelona, quedó demostrado que el método del Subgradiente no es el apropiado para resolver el problema de programación matemática, debido a que el método depende demasiado del punto inicial y por lo tanto no es estable. De acuerdo a lo propuesto en [18], para solucionar esta situación sería necesario implementar un método de regularización.

Adicionalmente, el modelo fue resuelto mediante tres variaciones del Cutting Plane, de las cuales dos consisten en acelerar el algoritmo, donde solo el método con MSWA lo logra, reduciendo la cantidad de iteraciones en cerca de un 30 % para el caso de Madrid y en un 45 % para el caso de Barcelona.

Asimismo, se demostró que el método de aceleración propuesto realmente ralentiza el algoritmo de Cutting Plane, debido a que el proceso de eliminar cortes en ciertas iteraciones lo desestabiliza, resultando en la generación de más iteraciones para lograr converger.

Con respecto a los resultados numéricos, pese a que en términos de costos y porcentaje de viajes a pie son relativamente apropiados, en la realidad no son prácticos de implementar, dado que para ambos casos el modelo selecciona muchas líneas con muy poca cantidad de vehículos y servicios, y también muchas de ellas comparten los mismos (o similares) recorridos, provocando que los headway sean altos, los tiempos de espera en la cola de pasajeros para abordar un autobús no sean convenientes y la cantidad de servicios que atienden las paradas alcancen el límite máximo. Esta situación se debe al hecho de que la recuperación de la solución primal considera todas las soluciones obtenidas en el proceso iterativo.

A partir de este trabajo se presentan oportunidades de mejora para futuras investigaciones:

- Con el fin de tener soluciones más prácticas, podría ser conveniente limitar la cantidad mínima de vehículos y/o servicios a asignar a cada una de las líneas. Así como también sería interesante buscar un método que seleccione solo aquellos cortes que realmente hayan mejorado la solución para realizar la recuperación de la solución primal.
- Aplicar el método a redes más grandes.
- Considerar situaciones de elevada congestión.
- Considerar la asignación de pasajeros bajo el concepto de estrategias.

Apéndice A

Valor de los coeficientes $\tilde{\beta}$ y $\tilde{\gamma}$

Tabla 8. coeficientes $\tilde{\beta}$ y $\tilde{\gamma}$ de la aproximación (61) para los modelos de cola $M/M^{[Y]}/1$. La capacidad residual c del bus se distribuye siguiendo una normal truncada en $[0,100]$ con un valor promedio de 50 pasajeros. Los valores son reportados para varios coeficientes de variación para c , C_c , en el rango de 0.0 a 0.5.

ρ	$C_c = 0.04$	0.1	0.2	0.3	0.4	0.5
$\tilde{\beta}$						
0.05	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008
0.15	0.0015	0.0016	0.0026	0.0043	0.0073	0.0117
0.25	0.0225	0.0235	0.0270	0.0328	0.0405	0.0508
0.35	0.0952	0.0973	0.1043	0.1158	0.1312	0.1508
0.45	0.2709	0.2750	0.2880	0.3096	0.3388	0.3766
0.55	0.6694	0.6771	0.7024	0.7447	0.8036	0.8806
0.65	1.6263	1.6426	1.6972	1.7875	1.9156	2.0878
0.75	4.3380	4.3777	4.5128	4.7385	5.0655	5.5178
0.85	15.4343	15.5703	16.0344	16.8138	17.9800	19.6775
0.95	160.2669	161.6596	166.4096	174.5540	188.2396	212.0599
$\tilde{\gamma}$						
0.05	1.0000	1.0000	1.0000	1.0020	1.0110	1.0310
0.15	1.0110	1.0130	1.0200	1.0380	1.0740	1.1340
0.25	1.1110	1.1170	1.1370	1.1750	1.2350	1.3250
0.35	1.3490	1.3590	1.3910	1.4480	1.5330	1.6540
0.45	1.7830	1.7970	1.8450	1.9270	2.0470	2.2120
0.55	2.5720	2.5930	2.6650	2.7880	2.9670	3.2090
0.65	4.1500	4.1860	4.3060	4.5080	4.8010	5.2010
0.75	7.9790	8.0470	8.2810	8.6740	9.2480	10.0430
0.85	21.6390	21.8260	22.4650	23.5400	25.1460	27.4730
0.95	178.3520	179.8980	185.1750	194.2170	209.3520	235.5510

Apéndice B

Matrices origen-destino de Madrid y Barcelona

Tabla 9. Matriz de viajes origen-destino (estación a estación) durante un periodo de $H = 180$ minutos para el corredor ferroviario de Madrid. La última fila y columna corresponden a las tasas promedios de llegadas y salidas por minuto en la estación

	At	Re	NM	Ch	Total Or.	pax/min
At	0	2011	22097	368	24476	135.98
Re	170	0	3066	230	3466	19.25
NM	4386	150	0	170	4706	26.14
Ch	2504	150	2438	0	5092	28.28
Total Dest.	7060	2311	27601	768	37740	—
pax/min	39.22	12.84	153.34	4.26	—	—

Tabla 10. Matriz de viajes origen-destino (estación a estación) durante un periodo de $H = 180$ minutos para la línea 1 del metro de Barcelona. La última fila y columna corresponden a las tasas promedios de llegadas y salidas por minuto en la estación

	AT	Cl	Ma	PC	PE	PG0	PQ	PU	Ro	Ur	Total Or.	pax/min
AT	0	1764	10	1152	1476	720	10	576	144	144	5996	33.31
Cl	864	0	360	960	1230	600	576	480	120	120	5310	29.50
Ma	432	735	0	480	615	10	288	240	60	60	2920	16.22
PC	1152	1960	480	0	1640	800	10	10	160	0	6212	34.51
PE	576	980	240	640	0	400	384	320	80	80	3700	20.56
PG0	720	1225	10	800	1025	0	480	400	100	100	4860	27.00
PQ	10	1176	288	10	984	480	0	384	96	96	3524	19.58
PU	576	980	240	10	820	400	384	0	80	10	3500	19.44
Ro	144	245	60	160	205	100	96	80	0	10	1100	6.11
Ur	144	245	60	0	205	100	96	10	10	0	870	4.83
Total Dest.	4618	9310	1748	4212	8200	3610	2324	2500	850	620	37992	—
pax/min	25.66	51.72	9.71	23.40	45.56	20.06	12.91	13.89	4.72	3.44	—	—

Bibliografía

- [1] O. Ibarra-Rojas, F. Delgado, R. Giesen, and J. Muñoz, “Planning, operation, and control of bus transport systems: A literature review,” in *Transportation Research Part B Methodological*, vol. 77, pp. 38–75, Elsevier, 12 2014.
- [2] G. Desaulniers and M. D. Hickman, “Public transit,” in *Transportation* (C. Barnhart and G. Laporte, eds.), vol. 14 of *Handbooks in Operations Research and Management Science*, pp. 69 – 127, Elsevier, 2007.
- [3] A. Mautone, *Optimización de Recorridos y Frecuencias en Sistemas de Transporte Público Urbano Colectivo*. PhD thesis, Departamento de Investigación Operativa. Universidad de La República, Julio 2005.
- [4] A. Ceder and N. H. Wilson, “Bus network design,” *Transportation Research Part B: Methodological*, vol. 20, no. 4, pp. 331 – 344, 1986.
- [5] M. Baaj and H. Mahmassani, “An ai-based approach for transit route system planning and design,” *Journal of Advanced Transportation*, vol. 25, pp. 187–209, 1 1991.
- [6] Y. Israeli and A. Ceder, “Transit route design using scheduling and multiobjective programming techniques,” in *Computer-Aided Transit Scheduling*, (Berlin, Heidelberg), pp. 56–75, Springer Berlin Heidelberg, 1995.
- [7] V. M. Tom and S. Mohan, “Transit route network design using frequency coded genetic algorithm,” in *Journal of Transportation Engineering-asce - J TRANSP ENG-ASCE*, vol. 129, 03 2003.
- [8] E. Codina, A. Marín, and F. López, “A model for setting services on auxiliary bus lines under congestion,” *TOP*, vol. 21, pp. 48–83, Apr 2013.
- [9] H. Eiselt and C. Sandblom, *Integer Programming and Network Models*. Springer, 2000.
- [10] A. M. Geoffrion, *Lagrangian relaxation for integer programming*, pp. 82–114. Berlin, Heidelberg: Springer Berlin Heidelberg, 1974.
- [11] M. L. Fisher, “The lagrangian relaxation method for solving integer programming problem,” *Management Science*, vol. 27, pp. 1–18, 01 1981.
- [12] M. L. Fisher, “An applications oriented guide to lagrangian relaxation,” *Interfaces*, vol. 15, pp. 10–21, 04 1985.
- [13] J. E. Beasley, “Lagrangian relaxation,” in *Modern Heuristic Techniques for Combinatorial Problems* (C. R. Reeves, ed.), pp. 243–303, New York, NY, USA: John Wiley & Sons, Inc., 1993.
- [14] H. Liu, X. S. He, and B.-S. He, “Method of successive weighted averages (mswa) and self-regulated averaging schemes for solving stochastic user equilibrium problem,” in *Networks and Spatial Economics*, vol. 9, pp. 485–503, 12 2009.
- [15] K. M. Anstreicher and L. A. Wolsey, “Two well-known properties of subgradient optimization,” *Math. Program.*, vol. 120, pp. 213–220, Apr. 2009.
- [16] M. Minoux, *Mathematical Programming: Theory and Algorithms*. Wiley-Interscience series in discrete mathematics and optimization, Wiley, 1986.
- [17] S. Lawphongpanich and D. Hearn, “Cutting plane algorithms for maximin problems,” 1991.
- [18] B. Polyak, “Subgradient methods: a survey of soviet research,” in *Nonsmooth Optimization*, Pergamon Press, 01 1977.